

**T-CELL RECEPTOR $\gamma$  ALTERNATE READING FRAME PROTEIN,  
(TARP) AND USES THEREOF**

**CROSS-REFERENCES TO RELATED APPLICATIONS**

5           This application claims priority from U.S. Provisional Application No. 60/143,560, filed July 13, 1999, and U.S. Provisional Application No. 60/157,471, filed October 1, 1999. The contents of both of these applications are hereby incorporated by reference.

**STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER  
FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT**

10

Not applicable.

**BACKGROUND OF THE INVENTION**

This invention is directed to the fields of molecular biology and medical diagnostics and therapeutics.

15

Prostate cancer is the most prevalent form of human cancer and the third most common cause of cancer death in men. Methods for early detection and treatment of this disease would decrease the rate of prostate cancer deaths. Tumor-associated proteins, which are proteins expressed by malignant cells but few others, are useful as targets for detection and for intervention. Several proteins associated with prostate cancer have been identified, including prostate specific antigen (PSA).

20

Immunotherapy is a potent new weapon against cancer. Immunotherapy involves evoking an immune response against cancer cells based on their production of target antigens. While humoral immune responses against cancer cell antigens have uses, it is preferred to invoke a cell-mediated immune response against cancer cells.

25

Immunotherapy based on cell-mediated immune responses involves generating a cell-mediated response to cells that produce particular antigenic determinants. Cancer cells produce various proteins that can become the target of immunotherapy. Certain cancers produce novel proteins, for example as a result of mutation, that are immunogenic.

30

However, investigators also have discovered tumor infiltrating lymphocytes that specifically recognize un-mutated proteins of cancer cells. For example, Rosenberg *et al.* have shown that tumor infiltrating lymphocytes target and recognize antigenic

determinants of the protein MART-1, produced by both normal melanocytes and malignant melanoma cells. Furthermore, active or passive immunotherapy directed against MART-1 or peptides of it that bind to MHC Class I molecules (epitopes of HLA A2, in particular) results in the destruction of melanoma cells as well normal cells that  
5 produce MART-1. Y. Kawakami *et al.*, *J. Immunol.* 21:237 (1998).

Novel cancer antigens are expected to provoke an immune response because the immune system will recognize them as non-self proteins. However, the ability of the immune system to invoke an immune response against an un-mutated self protein was surprising because the immune system develops tolerance to self proteins. It  
10 is believed that this immune response is directed against antigenic determinants that normally are not exposed to the immune system in sufficient quantity to invoke either tolerance or an immune response. In cancer, however, these determinants no longer escape detection by the immune system. This may result from increased presentation of the determinants by MHC Class I molecules.

15 The cell-mediated immune response involves the activity of Major Histocompatibility Complex molecules. In humans, this complex is called the "HLA" ("Human Leukocyte Antigen") complex. In mice, it is referred to as the "H-2" complex. The major histocompatibility complex includes three classes of proteins, MHC class I, MHC class II and MHC class III. MHC class I molecules are expressed on the surface of  
20 nearly all nucleated cells. They present antigen peptides to T<sub>C</sub> cells (CD8+). There are three MHC class I gene loci in humans, HLA A, HLA B and HLA C. Each locus is highly polymorphic. Therefore, a person may have up to six different kinds of HLA molecules on the surface of their cells. MHC Class II proteins are expressed primarily on antigen presenting cells such as macrophages, dendritic cells and B cells, where they  
25 present processed antigenic peptides to T<sub>H</sub> cells. There are three MHC Class II gene loci in humans, HLA DP, HLA DQ and HLA DR. MHC class III proteins are associated with various immune processes, and include soluble serum proteins, components of the complement system and tumor necrosis factors. J. Kuby, Chapter 9, IMMUNOLOGY, Third Edition W.H. Freeman and Company, New York (1997).

30 In cancer cells as well as healthy cells, MHC class I molecules present epitopes from endogenous proteins for presentation to T<sub>C</sub> cells. HLA A, HLA B and HLA C molecules bind peptides of about 8 to 10 amino acids in length that have particular anchoring residues. The anchoring residues recognized by an HLA class I molecule depend upon the particular allelic form of the HLA molecule. A CD8+ T cell

bears T cell receptors that recognize a specific epitope when presented by a particular HLA molecule on a cell. When a T<sub>C</sub> cell that has been stimulated by an antigen presenting cell to become a cytotoxic T lymphocyte contacts a cell that bears such an HLA-peptide complex, the CTL forms a conjugate with the cell and destroys it.

5                   The presentation of peptides by MHC Class I molecules involves the cleavage of an endogenously produced protein into peptides by the proteasome, its processing through the ER and Golgi apparatus, its binding to the cleft in an MHC Class I molecule through the anchor residues of the peptide and ultimate presentation on the cell surface. Depending upon the particular anchor residues, among other things, certain  
10 peptides may bind more tightly to a particular HLA molecules than others. Peptides that bind well are referred to as "dominant" epitopes, while those that bind less well are termed "subdominant" or "cryptic" epitopes. Dominant epitopes of either self proteins or foreign proteins evoke strong tolerance or immune responses. Subdominant or cryptic epitopes generate weak responses or no responses at all. It is hypothesized that tighter  
15 binding by dominant epitopes to HLA molecules results in their denser presentation on the cell surface, greater opportunity to react with immune cells and greater likelihood of eliciting an immune response or tolerance.

Investigation has shown that in the case of the MART-1 protein, a self protein, the immune system generates the greatest CTL response against subdominant or  
20 cryptic epitopes. Y. Kawakami *et al.* 1997 *Immunol. Res.* 16:313. It may be that in cancer cells subdominant or cryptic epitopes are presented much more densely or in greater amounts than is normal; consequently, the immune system encounters these previously undetected epitopes, recognizes them as foreign and generates an immune response against them. Whatever the reason, exposing the immune system to large  
25 amounts of subdominant or cryptic epitopes of self proteins as a means of eliciting an immune response against cells that produce that protein is a key element of cancer immunotherapy. Of course, eliciting an immune response against a self protein will result in the destruction of both cancerous cells and healthy cells. Therefore, for such  
immunotherapy to succeed, the healthy cells must either be non-essential for life or have  
30 functions that are replaceable by other therapies. In the case of prostate cancers and breast cancer, surgical removal of the prostate or breast, respectively, is a frequent therapeutic intervention. In such cases, most or all of the tissue displaying a prostate or breast antigen will be a tumor cell.

## SUMMARY OF THE INVENTION

We have discovered that prostate cells of epithelial origin both healthy and cancerous, transcribe a portion of an unrearranged TCR $\gamma$  gene. While *in vitro*, this transcript resulted in two proteins, one of which is a truncated form of TCR $\gamma$ , our studies show, surprisingly, that *in vivo*, the protein expressed is not a TCR $\gamma$  protein, but a protein expressed from an alternate reading frame. Accordingly, the protein has been designated the TCR $\gamma$  Alternate Reading Frame Protein, or "TARP."

We have discovered that TARP is expressed in prostate cells of epithelial origin and in prostate cancer cells. Surprisingly, we have discovered that the same protein is expressed in many breast cancer cells. TARP is therefore useful as a marker of prostate cancer cells and of breast cancer cells which express TARP ("TARP-expressing breast cancers") and as a basis for immunotherapy. This invention provides both the nucleic acids and the protein in isolated or recombinant form. Although we have now found that the protein is expressed in many breast cancer cells, it was first identified in prostate cells, and will sometimes be referred to below as prostate-specific ("PS")-TCR $\gamma$ .

In one aspect, the invention provides an isolated polypeptide comprising the amino acid sequence of the TCR $\gamma$  Alternate Reading frame Protein ("TARP"). The invention further provides immunogenic fragments of TARP (fragments which can raise an antibody which specifically recognizes and binds to full-length TARP or which activates a T-cell to recognize a cell expressing TARP), polypeptides with at least 90% sequence identity to TARP and which are specifically recognized by antibodies which specifically recognizes TARP, and polypeptides with at least 90 % sequence identity with TARP and which, when processed and presented in the context of Major Histocompatibility Complex molecules, activate T lymphocytes against cells which express TARP. The invention further provides compositions in TARP, immunogenic fragments thereof, or peptides with at least 90% sequence identity and which meet the functional criteria noted above are present in a pharmaceutically acceptable carrier.

In another set of embodiments, the invention provides isolated, recombinant nucleic acid molecules which encode a polypeptide having the amino acid sequence of a TCR $\gamma$  Alternate Reading frame Protein ("TARP"), which encode an immunogenic fragment thereof, which encode a polypeptide with at least 90% sequence identity to TARP and which is specifically recognized by an antibody which specifically recognizes TARP, or which encode a polypeptide which has at least 90 % sequence

identity with TARP and which, when processed and presented in the context of Major Histocompatibility Complex molecules, activates T lymphocytes against cells which express TARP.

In yet another series of embodiments, the invention provides methods comprising administering to a subject a composition, which composition is selected from the group consisting of: an isolated polypeptide having the amino acid sequence of a TCR $\gamma$  Alternate Reading frame Protein ("TARP"), an immunogenic fragment thereof, a polypeptide with at least 90% sequence identity to TARP and which is specifically recognized by an antibody which specifically recognizes TARP, and a polypeptide which has at least 90 % sequence identity with TARP and which, when processed and presented in the context of Major Histocompatibility Complex molecules, activates T lymphocytes against cells which express TARP.

The invention further provides methods of administering to a subject a composition, which composition is selected from the group consisting of: an isolated nucleic acid encoding TARP, an immunogenic fragment thereof, a polypeptide with at least 90% sequence identity to TARP and which is specifically recognized by an antibody which specifically recognizes TARP, and a polypeptide which has at least 90 % sequence identity with TARP and which, when processed and presented in the context of Major Histocompatibility Complex molecules, activates T lymphocytes against cells which express TARP.

The invention further provides methods of administering to a subject a composition, which composition comprises an antigen presenting cell pulsed with a polypeptide comprising an epitope of TARP. Additionally, the invention provides methods of administering to a subject a composition, which composition comprises cells sensitized in vitro to TARP, an immunogenic fragment thereof, a polypeptide with at least 90% sequence identity to TARP which is specifically recognized by an antibody which specifically recognizes TARP, or a polypeptide which has at least 90 % sequence identity with TARP which, when processed and presented in the context of Major Histocompatibility Complex molecules, activates T lymphocytes against cells which express TARP.

The compositions of the methods described above can be administered to a subject who suffers from prostate cancer, to a subject who suffers from breast cancer, or to a female subject who has not been diagnosed with breast cancer.

The methods further contemplate sensitizing CD8+ cells *in vitro* to an epitope of a TARP protein and administering the sensitized cells to the subject. The CD8+ cells may be T<sub>C</sub> cells. In some embodiments, the T<sub>C</sub> cells may be are tumor infiltrating lymphocytes.

5                    Additionally, the methods may comprise co-administering to the subject an immune adjuvant selected from non-specific immune adjuvants, subcellular microbial products and fractions, haptens, immunogenic proteins, immunomodulators, interferons, thymic hormones and colony stimulating factors.

10                    The methods may further comprise administering an antigen presenting cell pulsed with a polypeptide comprising an epitope of TARP or administering a nucleic acid sequence encoding polypeptide comprising an epitope of TARP, which nucleic acid is in a recombinant virus. The methods further comprise administering a nucleic acid sequence encoding a polypeptide comprising an epitope of a TARP protein.

15                    Additionally, the methods comprise administering an expression vector that expresses a polypeptide comprising an epitope of a TARP protein, which expression vector is in a recombinant bacterial cell. Further, the methods comprise immunizing a subject with a expression vector that expresses a polypeptide comprising an epitope of a TARP protein, which expression vector is in an autologous recombinant cell.

20                    In another aspect, this invention provides a method for detecting, in a male, a prostate cell of epithelial origin, or, in a female, a breast cancer cell, comprising detecting in a cell from said male or said female a nucleic acid transcript encoding TARP, or detecting TARP produced by translation of the transcript, whereby detection of the transcript or of the protein in a cell from said male identifies the cell as a prostate epithelial cell and whereby detection of the transcript or of the protein in a cell from said  
25                    female identifies the cell as a breast cancer cell. The methods may comprise contacting RNA from the cell with a nucleic acid probe that specifically hybridizes to the transcript under hybridization conditions, and detecting hybridization. Moreover, the methods may include disrupting a cell and contacting a portion of the cell contents with a chimeric molecule comprising a targeting moiety and a detectable label, wherein the targeting  
30                    moiety specifically binds to the protein, and detecting the label bound to the protein. The targeting moiety itself may also be labelled (for example, an antibody such as an scFv may have a radioactive residue).

The cell being examined may be from a lymph node, or it may be from a breast biopsy.

Finally, the invention provides antibodies that specifically bind to an epitope of a TCR $\gamma$  Alternate Reading frame Protein.

### BRIEF DESCRIPTION OF THE DRAWINGS

5

**FIGURE 1.** Nucleotide sequence (SEQ ID NO:13) of PS-TCR $\gamma$  transcript.

In *in vitro* translation systems, this transcript produces two polypeptides. A first polypeptide has a deduced amino acid sequence beginning MQM . . . ("PS-TCR $\gamma$ -1", now called "TARP," SEQ ID NO:14). This polypeptide has a predicted mass of 7.2 kD. It is translated from a reading frame which does not coincide with the natural TCR $\gamma$  reading frame. A second polypeptide has a deduced amino acid sequence beginning MKT . . . ("PS-TCR $\gamma$ -2" SEQ ID NO:15). It has a predicted mass of 13 kD. It is translated from the same reading frame as TCR $\gamma$  and represents a truncated version of TCR $\gamma$ . Single underlined sequences in the figure indicate the transcription initiation site and a polyadenylation site.

10  
15

**FIGURE 2.** Hybridization analysis of TCR $\gamma$  mRNA expression. Figure 2A) Multiple tissue dot blot showing differential expression of human TCR $\gamma$ . Positive tissues are prostate (C7), small intestine (E3), spleen (E4), thymus (E5), peripheral leukocyte (E6), lymph node (E7), bone marrow (E8) and lung (F2). Figure 2B). Northern blot showing TCR $\gamma$  transcript sizes in normal tissues. Two TCR $\gamma$  transcripts expressed in prostate are 1.1 kb and 2.8 kb while the predominant transcript in spleen, thymus and peripheral blood leukocytes is 1.5 kb. The film was exposed for 20 hours.

20

**FIGURE 3.** Northern blot analysis of TCR  $\gamma\delta$  expression. Figure 3A) A TCR $\gamma$  constant domain (TCR C $\gamma$ ) cDNA probe shows the 1.1 and 2.8 kb prostate-specific transcripts (compare with Fig. 2B). The film was exposed for 20 hours. Figure 3B) A TCR $\delta$  constant domain (TCR C $\delta$ ) cDNA probe reveals that TCR $\delta$  mRNA is not expressed in prostate while expression is seen in spleen, thymus and peripheral blood leukocytes. The film was exposed for 50 hours. Figure 3C) A TCR C $\gamma$  cDNA probe shows that the LNCaP cell line expresses TCR $\gamma$  while the PC-3 cell line does not. The film was exposed for 20 hours. Human  $\beta$ -actin mRNA expression was analyzed as a control.

25  
30

**FIGURE 4.** RNA in situ hybridization on paraffin-embedded tissue sections using a TCR $\gamma$  (C $\gamma$ 1-3' UTR) anti-sense, <sup>35</sup>S-labeled riboprobe. The left panel photos are from dark field microscopy, while the corresponding right panel photos are

from in bright field microscopy. The bright grains shown in pictures taken in dark field are signals of RNA hybridization. 4A) Prostate tissues from a 67 year old man showing positive acinar epithelial cells and negative stromal cells, (5x magnification). 4B) Bright field of 4A. 4C) Higher magnification (40x) showing positive areas in the lower right corner. 4D) Bright field of 4C. 4E) Kidney tissues showing no RNA hybridization, (5x magnification). 4F) Bright field of 4E.

**FIGURE 5.** Primer-extension of LNCaP mRNA. The reverse primer anneals in the constant domain of TCR $\gamma$ , starting 75 nucleotides from the 5' end of C $\gamma$ 1. The reverse transcription stopped at approximately 128 nucleotides, indicated by the arrow, revealing that the transcript is initiated approximately 53 nucleotides upstream of C $\gamma$ 1. The lane with TCR $\gamma$  reverse transcription of LNCaP was exposed for 72 hours while the marker lane was exposed for 8 hours.

**FIGURE 6.** The prostate TCR $\gamma$  transcript. Illustration on how the prostate TCR $\gamma$  is transcribed and spliced. The transcript consist of a J $\gamma$ 1.2 segment, the three exons of C $\gamma$ 1, followed by untranslated sequence.

**FIGURE 7.** *In vitro* transcription-coupled translation analysis of the prostate TCR $\gamma$ . Two proteins with estimated sizes of 8 and 13 kDa were obtained (lane 1). Negative control reactions using the empty vector (lane 2) did not yield any protein product.

**FIGURE 8.** Primers used for analysis of PS-TCR $\gamma$  transcript. (SEQ ID NOs:1-12).

**FIGURE 9.** *In vitro* translation analysis of the prostate-specific TCR $\gamma$  transcript. The prostate-specific TCR $\gamma$  transcript encodes two proteins *in vitro*. <sup>35</sup>S-Met labeled *in vitro* translated proteins were run on a 16.5% Tris-Tricine gel and analyzed by autoradiography. A schematic representation of the mutant constructs used is shown on the right. An open box represents the first reading frame with potential initiation codons in bold whereas the second reading frame is represented by a shaded box with the potential initiation codon in italics. "X" indicates an ATG codon mutated to ATA. Size markers in kDa are indicated on the top.

**FIGURE 10.** TARP is nuclear protein expressed in prostate extracts. Western blot of protein extracts derived from LNCaP cells, PC3 cells or a prostate tumor sample (Cancer). 20  $\mu$ g of each protein extract were run on a 16.5% Tris-Tricine gel and probed with an antibody against TARP (top panel) or TCR $\gamma$  (bottom panel). As a positive



control, 1  $\mu$ g of His-tagged TARP (top panel) or 100 ng of His-tagged TCR $\gamma$  (bottom panel) were run on the gel (Recomb.). Size markers in kDa are indicated on the left. (B) Western blot of the cytoplasmic fraction (Cytoplasm), membrane fraction (Membrane) and nuclear fraction (Nucleus,) of LNCaP cells. 40  $\mu$ g of each fraction were run on a 16.5% Tris-Tricine gel and probed with an antibody against TARP. As a positive control, 1  $\mu$ g of His-tagged TARP was run on the gel (Recomb.). Size markers in kDa are indicated on the left.

**FIGURE 11. *TARP* mRNA is expressed in breast cancer cells.**

(A) RT-PCR was performed with primers specific for *TARP* (top panels) or *actin* (bottom panels) using RNA derived from the following cell lines: prostate (LNCaP and PC3), neuroblastoma (A172), colon (COLO 205), gastric (KATO III) and breast (MCF7, BT-474, Hs57Bst, SK-BR-3, CRL-1897 and MDA-468). RT-PCR reactions performed without template are indicated as dH<sub>2</sub>O. (B) PCR was performed using cDNAs derived from 12 human breast cancer tissue samples (lanes 1-12) using primers specific for *TARP* (top panel) or *actin* (bottom panel). PCR reactions performed without template are indicated as dH<sub>2</sub>O. For both panels, 20% of the PCR products were run on a 1% agarose gel and visualized by ethidium bromide staining.

**FIGURE 12. The *TARP* transcript found in breast cell line is the same as the prostate-specific form. (A) Schematic of the *TCR $\gamma$*  locus and how *TARP* is transcribed and spliced in prostate cells. Primers used for RT-PCR analysis in Panel B are indicated. (B) RT-PCR analysis of *TARP* mRNA expression. PCR reactions using *TARP* primers 1 and 3 (top panel), *TARP* primers 2 and 3 (middle panel) or *actin* primers (bottom panel) were performed with cDNAs derived from prostate cell lines (LNCaP and PC3) and breast cell lines (MCF7, BT-474, SK-BR-3 and Hs578Bst). RT-PCR reactions performed without template are indicated as dH<sub>2</sub>O. 20% of the PCR products were run on a 1% agarose gel and visualized by ethidium bromide staining. (C) Northern blot analysis of *TARP* transcripts. 2  $\mu$ g poly(A) mRNA from prostate cell lines (LNCaP and PC3) and breast cell lines (MCF7, BT-474, SK-BR-3 and Hs578Bst) were analyzed using a constant domain fragment as probe. The autoradiograph was generated after a 24-hour exposure (top panel). The same filter was stripped and analyzed with a human  $\beta$ -*actin* RNA probe to verify equal loading. The autoradiograph was generated after a 1-hour exposure (bottom panel). RNA size markers in the nucleotides are indicated on the left.**

**FIGURE 13.** TARP exists in the nuclei of breast cancer cells. (A) Western blot of nuclear extracts derived from LNCaP, MCF7, BT-474, SK-BR-3 and Hs57BsT cells. 40 µg of each nuclear extract were run on a 16.5% Tris-Tricine gel and probed with an antibody against TARP (upper panel) or TCRγ (bottom panel). As a positive control, 1 µg of His-tagged TARP (His-TARP) and 100 ng of His-tagged TCRγ (His-TCRγ) were run on the gels. Size markers in kDa are indicated on the left.

**FIGURE 14.** Potential functional domains of TARP. (A) TARP contains a potential leucine zipper motif and phosphorylation sites. A potential leucine zipper motif is indicated with boxed leucines followed by a basic region that is underlined. cAMP- and cGMP-dependent protein kinase phosphorylation sites (amino acids 46-49 and 55-58) and protein kinase C phosphorylation sites (amino acids 19-21 and 20-22) are outlined. (B) Protein sequence comparison of TARP with Tup1. Amino acids sequences for TARP (42-57), *Dictyostelium discoideum* Tup1 (dTup1, 521-536) and *Saccharomyces cerevisiae* Tup1 (yTup1, 626-660) are shown. Conserved residues are boxed.

## DETAILED DESCRIPTION OF THE INVENTION

### I. INTRODUCTION

Surprisingly, it has been discovered that prostate cells of epithelial origin, and cells of many breast cancer s, express mRNA of the T-cell receptor gamma chain ("TCRγ"). The major TCRγ transcript in prostate has a different size than that expressed in T lymphocytes. The findings that prostate epithelial cells and many breast cancers express a high level of a transcript from a gene thought to be expressed exclusively in T lymphocytes is highly unexpected.

Because the TCRγ reading frame contains a good Kozak sequence (Kozak, M. *Cell* 44:283-92 (1986)), we initially hypothesized that a truncated TCRγ protein was encoded. Thus, it was an additional surprise to find that the *TCRγ* locus expressed in epithelial prostate cells and breast cancer cells encodes a 7 kDa nuclear protein. Because the protein is encoded from a reading frame different from TCRγ, we have named it "TARP," for TCRγ Alternate Reading frame Protein. Besides being translated from an alternate reading frame of a transcript originating within an intron of the *TCRγ* locus, TARP has two other unusual features. First, it is surprising to find such a small peptide in the cell because most are usually secreted. Second, TARP lacks a good Kozak sequence

***In vitro and in vivo uses***

The presence of this protein in prostate epithelial cells, prostate cancer cells, and cells of many breast cancers, creates a number of opportunities for *in vitro* and *in vivo* uses. First, antibodies raised against the protein can be used in *in vitro* assays to detect the presence of cells expressing TARP in a sample. For example, the Examples below demonstrate that TARP and the TARP transcript are not present, or are present at very low levels in normal breast cells, but are easily detected in cells of breast cancer which express TARP. Detection of high levels of TARP transcript or of TARP in breast cells removed from a subject therefore would be indicative of the presence of a TARP-expressing breast cancer in the subject. With respect to TARP expression in prostate cells, persons of skill in the art are aware that removal of the prostate is a frequent surgical intervention in aggressive prostate cancers. Detection of TARP in cells from an individual whose prostate has been removed is indicative that prostate cancer is present. (Persons of skill will recognize that approximately 1,000 men a year are diagnosed with breast cancer. Thus, it is not impossible that the patient also independently suffers from breast cancer. Given the fact that prostate cancer is over 200 times more common in men, and that the discussion concerns a patient already found to suffer from prostate cancer, the odds are small that the patient also independently suffers from breast cancer. The diagnosis can be confirmed by knowledge of the site from which the sample was taken, histologic and morphologic features of the cells, and other routine diagnostic criteria. In any event, since the presence of either condition demands further evaluation and monitoring, a determination that TARP-expressing cells are present in a male whose prostate has been removed is itself very useful regardless of whether the individual has prostate cancer, breast cancer, or both. Detection of TARP-expressing cells in a male who does not have prostate cancer is indicative of breast cancer.)

TARP itself, immunogenic fragments of TARP, and nucleic acids encoding TARP or immunogenic fragments thereof can also be used *in vitro* to activate cytotoxic T lymphocytes ("CTLs") derived from a subject to attack prostate cancer cells and TARP-expressing breast cancer cells when infused into the subject.

TARP itself, immunogenic fragments of TARP, nucleic acids encoding TARP, or immunogenic fragments thereof, can be administered to a subject, generally in a pharmaceutically acceptable carrier, to raise or to heighten an immune response to a prostate cancer or TARP-expressing breast cancer. Such compositions can be

administered therapeutically, in individuals who have been diagnosed as suffering from prostate cancer or a TARP-expressing breast cancer.

As discussed below, TARP is a nuclear protein which contains structural features characteristic of proteins which regulate transcription. It is expected that modulation of TARP levels in a cell will affect the growth of the cell, the replication of the cell, or both. Thus, modulation of TARP levels can be important in controlling a cancer cell's aggressiveness. TARP levels can be reduced in a cell by various modalities which impair the ability of the RNA transcript to be translated into a protein, such as ribozymes and antisense molecules. TARP levels in a cell can also be increased. For example, expression vectors containing nucleic acids encoding TARP, driven by a strong constitutive promoter, can be introduced into a cell. The constitutive transcription of the nucleic acids results in higher levels of protein expression in the cell.

The remainder of this section describes various structural features of TARP. The text continues with definitions used in this disclosure, and continues with discussions of the selection of immunogenic fragments of TARP, the administration of TARP to subjects, the formation of antibodies against TARP, detection of TARP transcript and protein, and pharmaceutical compositions.

#### **Structural features of TARP**

TARP contains five leucines in heptad repeats, suggesting that TARP contains a leucine zipper dimerization motif (Figure 7A). For this to be true, TARP must contain an amphipathic helix. One indication that TARP may contain an amphipathic helix is that serine and proline residues, residues believed to serve as a helix initiator, are found immediately before the first leucine repeat. Second, many charged amino acids are found within the heptad repeats thereby giving the helix an amphipathic nature and potentially serving as salt bridges with other helicies. Even though the presence of leucines in heptad repeats is a good indication of a leucine zipper motif, there are proteins identified containing five leucines in heptad repeats that are not considered leucine zipper proteins. For example, the crystal structures for karyopherin (Chook, Y. M. *et al.*, *Nature* 399:230-237 (1999)), *B. stearothermophilus* pyrimidine nucleoside phosphorylase (Pugmire, M. J. *et al.*, *Structure* 6:1467-1479 (1998)) and *T. thermophilus* phenylalanyl-tRNA synthetase (Mosyak, L. *et al.*, *Nat. Struct. Biol.* 2:537-547 (1995)) have shown that these proteins do not contain  $\alpha$ -helical structures in the region where the sequence contains five leucines in heptad repeats. Interaction and structure studies are needed to determine the significance of the leucine repeats found in TARP.

Another unusual feature of the TARP amino acid sequence is that a region of basic amino acids follows the potential leucine zipper motif (Figure 7A), suggesting a possible DNA-binding motif. However, the orientation of the basic region is rather unique in that it follows the leucine repeats rather than precedes them. Most leucine zipper proteins that bind DNA have the basic region before the leucine repeats (for a review, see (Chook, Y. M. *et al.*, *Nature* 399:230-237 (1999))). The basic region in TARP may only be functioning as a nuclear localization signal, but the fact that TARP is a nuclear protein strengthens the hypothesis that TARP may bind DNA.

To determine if TARP shares homology with any known proteins, we performed a protein BLAST search against GenBank. This search indicated that the amino acid sequence of TARP shares some homology to *Dictyostelium discoideum* Tup1 (GenBank accession no. AAC29438) and *Saccharomyces cerevisiae* Tup1 (Williams, F. E. *et al.*, *Mol. Cell. Biol.* 10:6500-6511 (1990)) (Figure 7C). Yeast Tup1 is normally found in a complex with Cyc8(Ssn6) and is required for transcriptional repression of genes that are regulated by glucose, oxygen and DNA damage (Tzamarias, D. *et al.*, *Genes Dev.* 9:821-831 (1995)). Neither Cyc8(Ssn6) nor Tup1 binds DNA, but each acts as a part of a corepressor complex through interactions with specific DNA-binding proteins such as  $\alpha 2$ , Mig1, Rox1 and a1 (Tzamarias, D. *et al.*, *Genes Dev.* 9:821-831 (1995)). The C'-terminal half of Tup1 contains six repeats of a 43-amino acid sequence rich in aspartate and tryptophan, known as WD-40 or  $\beta$ -transducin repeats (Williams, F. E. *et al.*, *Mol. Cell. Biol.* 10:6500-6511 (1990); Fong, H. K. *et al.*, *Proc. Natl. Acad. Sci. USA* 83:2162-2166 (1986)). WD-40 repeats have been identified in many proteins and play a role in protein-protein interactions. Importantly, Tup1 has been shown to interact with  $\alpha 2$  through two of its WD-40 repeats (Komachi, K. *et al.*, *Genes Dev.* 8:2857-2867 (1994)). It is interesting to note that TARP shares homology with the fifth WD-40 repeat of Tup1 (Figure 7C). Because TARP is a nuclear protein, its homology with Tup1 suggests that TARP is a member of a functional nuclear protein complex involved in transcriptional regulation.

The TARP antibody recognizes a doublet in prostate and breast nuclear extracts (Figure 6A). The faster 7 kDa band comigrates with the His-TARP recombinant protein, while the weaker band runs at a larger molecular weight. One possible explanation for the 9 kDa band is post-translational modifications. To determine if TARP contains any known post-translational modification sites, we analyzed the TARP amino acid sequence using the PROSITE program of the *Swiss Institute of Bioinformatics*

ExPASy proteomics server (<http://www.expasy.ch>) (Appel, R. D. *et al.*, *Trends Biochem. Sci.* 19:248-260 (1994); Hofmann, K. *et al.*, *Nucleic Acids Res.* 27:215-219 (1999)). As shown in Figure 7A, many potential phosphorylation sites were found including cAMP- and cGMP-dependent protein kinase phosphorylation sites (RRAT and RRGT) and protein kinase C phosphorylation sites (SSR and SRR). Phosphorylation has been shown in many cases to cause a protein to run at a larger apparent molecular weight on an SDS-PAGE gel. If this is the case, the results from Figure 6 may indicate that the unmodified form is prevalent in LNCaP cells and that only the phosphorylated form is present in MCF7 and SK-BR-3 cells. TARP may therefore be post-translationally modified when expressed in prostate and breast cancer cells.

Our initial studies of the *TARP* transcript did not reveal *TARP* expression in the breast (Essand, M. *et al.*, *Proc. Natl. Acad. Sci. USA* 96:9287-9292 (1999)). One possible explanation is that *TARP* is expressed at low levels in the normal breast and is difficult to detect. As described in the Results section, very weak signals were detected in a PCR analysis of normal breast samples as compared to the strong signals detected in the cancer samples. Therefore, the presence of *TARP* in breast cancer cells may indicate that *TARP* expression is induced after the oncogenic transformation of breast cells. In addition, the existence of *TARP* in breast cancer cells may indicate that *TARP* is regulated by estrogen. This hypothesis is strengthened by the identification of an element within the intronic promoter of *TARP* that combines an androgen response element (ARE) with an estrogen response element (ERE). This hybrid element consists of two half-sites specific to the ARE at the 5' end and to the ERE at the 3' end [(Zilliacus, J. *et al.*, *Mol. Endocrinol.* 9:389-400 (1995)) and unpublished data]]. Additional experiments are needed to determine if estrogen regulates *TARP*. There are instances, however, where mutant AREs cause the expression of certain prostate-specific genes in breast tumors. For example, *prostate specific antigen* (PSA) has been shown to be expressed in breast tumors (Majumdar, S. *et al.*, *Br. J. Cancer* 79:1594-1602 (1999)). Molecular analysis of the aberrant expression of PSA lead to the discovery of a single point mutation in one of the AREs found within the PSA promoter. It is believed that this mutation leads to the loss of androgen-regulated PSA expression in breast tumors (Majumdar, S. *et al.*, *Br. J. Cancer* 79:1594-1602 (1999)). It is unclear at this time whether a similar mutation in the *TARP* promoter occurs in the three breast cell lines tested.

The prostate is dependent on androgens for maintenance of its structure and function. When prostate cells become malignant, they often lose their androgen

dependence. In this study, we used two prostate cell lines that differ in their dependence on androgen for growth: LNCaP and PC3 cells. The androgen receptor is present in the androgen-dependent LNCaP cell line, but is absent in the androgen-independent PC3 cell line (Tilley, W. D. *et al.*, *Cancer Res.* 50:5382-5386 (1990)). As shown in Figure 3, *TARP* is expressed in LNCaP cells but not in PC3 cells. This result suggests that *TARP* expression may be regulated by androgen stimulation. The identification of an ARE-like element within the *TARP* promoter strengthens the idea that *TARP* is induced by androgens. Expression in LNCaP cell but not in PC3 cells suggests that *TARP* is important in regulating androgen-dependent responses.

## II. DEFINITIONS

Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. The following references provide a general definition of many of the terms used in this invention: Singleton *et al.*, *DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY* (2d ed. 1994); *THE CAMBRIDGE DICTIONARY OF SCIENCE AND TECHNOLOGY* (Walker ed., 1988); *THE GLOSSARY OF GENETICS*, 5TH ED., R. Rieger *et al.* (eds.), Springer Verlag (1991); and Hale & Marham, *THE HARPER COLLINS DICTIONARY OF BIOLOGY* (1991). As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

"T cell receptor" refers to a heterodimer found on the surface of T cells comprising an  $\alpha$  chain and a  $\beta$  chain or a  $\gamma$  and a  $\delta$  chain. T cell receptors recognize processed antigens associated with MHC molecules.

"T-cell receptor  $\gamma$  Alternate Reading frame Protein" and "TARP" refer to the polypeptide whose sequence is set forth, e.g., in Figure 14. The polypeptide is translated from a form of the T-cell receptor  $\gamma$  gene present as a transcript in prostate cells of epithelial origin, in prostate cancer cells, and in many breast cancers. Since "TARP" is an acronym the last part of which stands for the word "protein," "TARP protein" is redundant.

As used herein, nucleic acid transcripts from which TARP can be translated are referred to as "TARP nucleic acids" or "nucleic acids encoding TARP." The gene from which TARP is transcribed is referred to herein as a "TARP gene."

As used herein, an "immunogenic fragment" of TARP refers to a portion of TARP which, when presented by a cell in the context of a molecule of the Major Histocompatibility Complex, can in a T-cell activation assay, activate a T-lymphocyte against a cell expressing TARP. Typically, such fragments are 8 to 12 contiguous amino acids of TARP in length, although longer fragments may of course also be used.

In the context of comparing one polypeptide to another, "sequence identity is determined by comparing the sequence of TARP, as the reference sequence, to a test sequence. Typically, the two sequences are aligned for maximal or optimal alignment.

A "ligand" is a compound that specifically binds to a target molecule.

A "receptor" is compound that specifically binds to a ligand.

"Cytotoxic T lymphocytes" ("CTLs") are important in the immune response to tumor cells. CTLs recognize peptide epitopes in the context of HLA class I molecules that are expressed on the surface of almost all nucleated cells.

Tumor-specific helper T lymphocytes ("HTLs") are also known to be important for maintaining effective antitumor immunity. Their role in antitumor immunity has been demonstrated in animal models in which these cells not only serve to provide help for induction of CTL and antibody responses, but also provide effector functions, which are mediated by direct cell contact and also by secretion of lymphokines (e.g., IFN $\gamma$  and TNF- $\alpha$ ).

"Antibody" refers to a polypeptide ligand comprising at least a light chain or heavy chain immunoglobulin variable region which specifically recognizes and binds an epitope (e.g., an antigen). This includes intact immunoglobulins and the variants and portions of them well known in the art such as, Fab' fragments, F(ab)'<sub>2</sub> fragments, single chain Fv proteins ("scFv"), and disulfide stabilized Fv proteins ("dsFv"). An scFv protein is a fusion protein in which a light chain variable region of an immunoglobulin and a heavy chain variable region of an immunoglobulin are bound by a linker. Natural immunoglobulins are encoded by immunoglobulin genes. These include the *kappa* and *lambda* light chain constant region genes, the *alpha*,  $\gamma$ , *delta*, *epsilon* and *mu* heavy chain constant region genes, and the myriad immunoglobulin variable region genes. The term "antibody" includes polyclonal antibodies, monoclonal antibodies, chimeric antibodies and humanized antibodies, produced by immunization, from hybridomas, or recombinantly.



"Epitope" or "antigenic determinant" refers to a site on an antigen to which B and/or T cells respond. Epitopes can be formed both from contiguous amino acids or noncontiguous amino acids juxtaposed by tertiary folding of a protein. Epitopes formed from contiguous amino acids are typically retained on exposure to denaturing solvents whereas epitopes formed by tertiary folding are typically lost on treatment with denaturing solvents. An epitope typically includes at least 3, and more usually, at least 5 or 8-10 amino acids in a unique spatial conformation. Methods of determining spatial conformation of epitopes include, for example, x-ray crystallography and 2-dimensional nuclear magnetic resonance. See, e.g., *Epitope Mapping Protocols* in METHODS IN MOLECULAR BIOLOGY, Vol. 66, Glenn E. Morris, Ed (1996).

A ligand or a receptor "specifically binds to" a compound analyte when the ligand or receptor functions in a binding reaction which is determinative of the presence of the analyte in a sample of heterogeneous compounds. Thus, the ligand or receptor binds preferentially to a particular analyte and does not bind in a significant amount to other compounds present in the sample. For example, a polynucleotide specifically binds to an analyte polynucleotide comprising a complementary sequence and an antibody specifically binds under immunoassay conditions to an antigen analyte bearing an epitope against which the antibody was raised.

"Immunoassay" refers to a method of detecting an analyte in a sample in which specificity for the analyte is conferred by the specific binding between an antibody and a ligand. This includes detecting an antibody analyte through specific binding between the antibody and a ligand. See Harlow and Lane (1988) ANTIBODIES, A LABORATORY MANUAL, Cold Spring Harbor Publications, New York, for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity.

"Vaccine" refers to an agent or composition containing an agent effective to confer a therapeutic degree of immunity on an organism while causing only very low levels of morbidity or mortality. Methods of making vaccines are, of course, useful in the study of the immune system and in preventing and treating animal or human disease.

An "immunogenic amount" is an amount effective to elicit an immune response in a subject.

"Polypeptide" refers to a polymer composed of amino acid residues, related naturally occurring structural variants, and synthetic non-naturally occurring analogs thereof linked via peptide bonds, related naturally occurring structural variants,

and synthetic non-naturally occurring analogs thereof. Synthetic polypeptides can be synthesized, for example, using an automated polypeptide synthesizer. The term "protein" typically refers to large polypeptides. The term "peptide" typically refers to short polypeptides.

5                   Conventional notation is used herein to portray polypeptide sequences: the left-hand end of a polypeptide sequence is the amino-terminus; the right-hand end of a polypeptide sequence is the carboxyl-terminus.

                  "Fusion protein" refers to a polypeptide formed by the joining of two or more polypeptides through a peptide bond formed by the amino terminus of one  
10                   polypeptide and the carboxyl terminus of the other polypeptide. A fusion protein may be typically expressed as a single polypeptide from a nucleic acid sequence encoding the single contiguous fusion protein. However, a fusion protein can also be formed by the chemical coupling of the constituent polypeptides.

                  "Conservative substitution" refers to the substitution in a polypeptide of an  
15                   amino acid with a functionally similar amino acid. The following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Serine (S), Threonine (T);
- 2) Aspartic acid (D), Glutamic acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 20                   4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

                  Two proteins are "homologs" of each other if they exist in different species, are derived from a common genetic ancestor and share at least 70% amino acid  
25                   sequence identity.

                  "Substantially pure" or "isolated" means an object species is the predominant species present (*i.e.*, on a molar basis, more abundant than any other individual macromolecular species in the composition), and a substantially purified fraction is a composition wherein the object species comprises at least about 50% (on a  
30                   molar basis) of all macromolecular species present. Generally, a substantially pure composition means that about 80% to 90% or more of the macromolecular species present in the composition is the purified species of interest. The object species is purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods) if the composition consists essentially of

a single macromolecular species. Solvent species, small molecules (<500 Daltons), stabilizers (*e.g.*, BSA), and elemental ion species are not considered macromolecular species for purposes of this definition.

"Nucleic acid" refers to a polymer composed of nucleotide units  
5 (ribonucleotides, deoxyribonucleotides, related naturally occurring structural variants, and synthetic non-naturally occurring analogs thereof) linked via phosphodiester bonds, related naturally occurring structural variants, and synthetic non-naturally occurring analogs thereof. Thus, the term includes nucleotide polymers in which the nucleotides and the linkages between them include non-naturally occurring synthetic analogs, such as,  
10 for example and without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs), and the like. Such polynucleotides can be synthesized, for example, using an automated DNA synthesizer. The term "oligonucleotide" typically refers to short polynucleotides, generally no greater than about 50 nucleotides. It will be understood  
15 that when a nucleotide sequence is represented by a DNA sequence (*i.e.*, A, T, G, C), this also includes an RNA sequence (*i.e.*, A, U, G, C) in which "U" replaces "T."

Conventional notation is used herein to describe nucleotide sequences: the left-hand end of a single-stranded nucleotide sequence is the 5'-end; the left-hand direction of a double-stranded nucleotide sequence is referred to as the 5'-direction. The  
20 direction of 5' to 3' addition of nucleotides to nascent RNA transcripts is referred to as the transcription direction. The DNA strand having the same sequence as an mRNA is referred to as the "coding strand"; sequences on the DNA strand having the same sequence as an mRNA transcribed from that DNA and which are located 5' to the 5'-end of the RNA transcript are referred to as "upstream sequences"; sequences on the DNA  
25 strand having the same sequence as the RNA and which are 3' to the 3' end of the coding RNA transcript are referred to as "downstream sequences."

"cDNA" refers to a DNA that is complementary or identical to an mRNA, in either single stranded or double stranded form.

"Encoding" refers to the inherent property of specific sequences of  
30 nucleotides in a polynucleotide, such as a gene, a cDNA, or an mRNA, to serve as templates for synthesis of other polymers and macromolecules in biological processes having either a defined sequence of nucleotides (*i.e.*, rRNA, tRNA and mRNA) or a defined sequence of amino acids and the biological properties resulting therefrom. Thus, a gene encodes a protein if transcription and translation of mRNA produced by that gene

produces the protein in a cell or other biological system. Both the coding strand, the nucleotide sequence of which is identical to the mRNA sequence and is usually provided in sequence listings, and non-coding strand, used as the template for transcription, of a gene or cDNA can be referred to as encoding the protein or other product of that gene or cDNA. Unless otherwise specified, a "nucleotide sequence encoding an amino acid sequence" includes all nucleotide sequences that are degenerate versions of each other and that encode the same amino acid sequence. Nucleotide sequences that encode proteins and RNA may include introns.

"Recombinant nucleic acid" refers to a nucleic acid having nucleotide sequences that are not naturally joined together. This includes nucleic acid vectors comprising an amplified or assembled nucleic acid which can be used to transform a suitable host cell. A host cell that comprises the recombinant nucleic acid is referred to as a "recombinant host cell." The gene is then expressed in the recombinant host cell to produce, *e.g.*, a "recombinant polypeptide." A recombinant nucleic acid may serve a non-coding function (*e.g.*, promoter, origin of replication, ribosome-binding site, *etc.*) as well.

"Expression control sequence" refers to a nucleotide sequence in a polynucleotide that regulates the expression (transcription and/or translation) of a nucleotide sequence operatively linked thereto. "Operatively linked" refers to a functional relationship between two parts in which the activity of one part (*e.g.*, the ability to regulate transcription) results in an action on the other part (*e.g.*, transcription of the sequence). Expression control sequences can include, for example and without limitation, sequences of promoters (*e.g.*, inducible or constitutive), enhancers, transcription terminators, a start codon (*i.e.*, ATG), splicing signals for introns, and stop codons.

"Expression cassette" refers to a recombinant nucleic acid construct comprising an expression control sequence operatively linked to an expressible nucleotide sequence. An expression cassette generally comprises sufficient *cis*-acting elements for expression; other elements for expression can be supplied by the host cell or *in vitro* expression system.

"Expression vector" refers to a vector comprising an expression cassette. Expression vectors include all those known in the art, such as cosmids, plasmids (*e.g.*, naked or contained in liposomes) and viruses that incorporate the expression cassette.

A first sequence is an "antisense sequence" with respect to a second sequence if a polynucleotide whose sequence is the first sequence specifically hybridizes with a polynucleotide whose sequence is the second sequence.

Terms used to describe sequence relationships between two or more nucleotide sequences or amino acid sequences include "reference sequence," "selected from," "comparison window," "identical," "percentage of sequence identity," "substantially identical," "complementary," and "substantially complementary."

For sequence comparison of nucleic acid sequences, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters are used. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (*see, e.g., Current Protocols in Molecular Biology* (Ausubel *et al.*, eds 1995 supplement)).

One example of a useful algorithm is PILEUP. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35:351-360 (1987). The method used is similar to the method described by Higgins & Sharp, *CABIOS* 5:151-153 (1989). Using PILEUP, a reference sequence is compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps. PILEUP can be obtained from the GCG sequence analysis software package, *e.g.*, version 7.0 (Devereaux *et al.*, *Nuc. Acids Res.* 12:387-395 (1984)).

Another example of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and the BLAST 2.0 algorithm, which are described in Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990) and Altschul *et al.*, *Nucleic Acids Res.* 25:3389-3402 (1997)). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information

(<http://www.ncbi.nlm.nih.gov/>). The BLASTN program (for nucleotide sequences) uses as defaults a word length (W) of 11, alignments (B) of 50, expectation (E) of 10, M=5, N=4, and a comparison of both strands. The BLASTP program (for amino acid sequences) uses as defaults a word length (W) of 3, and expectation (E) of 10, and the  
5 BLOSUM62 scoring matrix (*see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)).

"Stringent hybridization conditions" refers to 50% formamide, 5 x SSC and 1% SDS incubated at 42° C or 5 x SSC and 1% SDS incubated at 65° C, with a wash in 0.2 x SSC and 0.1% SDS at 65° C.

10 "Naturally-occurring" as applied to an object refers to the fact that the object can be found in nature. For example, an amino acid or nucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally-occurring.

"Linker" refers to a molecule that joins two other molecules, either  
15 covalently, or through ionic, van der Waals or hydrogen bonds, *e.g.*, a nucleic acid molecule that hybridizes to one complementary sequence at the 5' end and to another complementary sequence at the 3' end, thus joining two non-complementary sequences.

"Pharmaceutical composition" refers to a composition suitable for pharmaceutical use in a mammal. A pharmaceutical composition comprises a  
20 pharmacologically effective amount of an active agent and a pharmaceutically acceptable carrier.

"Pharmacologically effective amount" refers to an amount of an agent effective to produce the intended pharmacological result.

"Pharmaceutically acceptable carrier" refers to any of the standard  
25 pharmaceutical carriers, buffers, and excipients, such as a phosphate buffered saline solution, 5% aqueous solution of dextrose, and emulsions, such as an oil/water or water/oil emulsion, and various types of wetting agents and/or adjuvants. Suitable pharmaceutical carriers and formulations are described in REMINGTON'S PHARMACEUTICAL SCIENCES, 19th Ed. (Mack Publishing Co., Easton, 1995). Preferred  
30 pharmaceutical carriers depend upon the intended mode of administration of the active agent. Typical modes of administration include enteral (*e.g.*, oral) or parenteral (*e.g.*, subcutaneous, intramuscular, intravenous or intraperitoneal injection; or topical, transdermal, or transmucosal administration). A "pharmaceutically acceptable salt" is a salt that can be formulated into a compound for pharmaceutical use including, *e.g.*, metal

salts (sodium, potassium, magnesium, calcium, *etc.*) and salts of ammonia or organic amines.

A "subject" of diagnosis or treatment is a human or non-human mammal.

5 "Administration" of a composition refers to introducing the composition into the subject by a chosen route of administration. For example, if the chosen route is intravenous, the composition is administered by introducing the composition into a vein of the subject.

"Treatment" refers to prophylactic treatment or therapeutic treatment.

10 A "prophylactic" treatment is a treatment administered to a subject who does not exhibit signs of a disease or exhibits only early signs for the purpose of decreasing the risk of developing pathology.

A "therapeutic" treatment is a treatment administered to a subject who exhibits signs of pathology for the purpose of diminishing or eliminating those signs.

15 "Diagnostic" means identifying the presence or nature of a pathologic condition. Diagnostic methods differ in their sensitivity and specificity. The "sensitivity" of a diagnostic assay is the percentage of diseased individuals who test positive (percent of true positives). The "specificity" of a diagnostic assay is 1 minus the false positive rate, where the false positive rate is defined as the proportion of those without the disease who test positive. While a particular diagnostic method may not provide a definitive  
20 diagnosis of a condition, it suffices if the method provides a positive indication that aids in diagnosis.

"Prognostic" means predicting the probable development (*e.g.*, severity) of a pathologic condition.

### III. TARP

25 This invention provides isolated, recombinant TARP. Because we first found isolated a prostate-specific TCR $\gamma$  transcript, we initially used the terms "PS-TCR $\gamma$  protein" and "PS-TCR $\gamma$  polypeptide" to refer to any polypeptide that could be translated in any reading frame from the ~1.1 kb PS-TCR $\gamma$  transcript. In particular, the terms referred to two proteins, PS-TCR $\gamma$ -1 (SEQ ID NO:14) and PS-TCR $\gamma$ -2 (SEQ ID NO:15),  
30 translated in *in vitro* translation systems. We have now determined that only the first of these reading frames is translated in prostate cells. Since this reading frame is not the reading frame which results in the TCR $\gamma$  chain, the protein is now referred to as the "T-

cell receptor Alternate Reading frame Protein.” Full-length TARP is a 58 amino acid protein whose sequence is set forth in SEQ ID NO:14 and Figure 14.

In certain embodiments, this invention provides polypeptides comprising an epitope comprising at least 5 to at least 15 consecutive amino acids from TARP. Such proteins bind to antibodies raised against full-length TARP (in this section, references to “TARP” refer to the full-length protein unless otherwise required by context). In other embodiments, this invention provides fusion proteins comprising a first and second polypeptide moiety in which one of the protein moieties comprises an amino acid sequence of at least 5 amino acids identifying an epitope of TARP. In one embodiment the TARP moiety is all or substantially of TARP. The other moiety can be, *e.g.*, an immunogenic protein. Such fusions also are useful to evoke an immune response against TARP. In other embodiments this invention provides TARP-like peptides (“TARP analogs”) whose amino acid sequences are at least 90% identical to TARP (although they may have 91%, 92%, 93%, 94%, 95%, or even higher sequence identity to TARP) and which are specifically bound by antibodies which specifically bind to TARP. In yet other embodiments this invention provides TARP-like peptides (also sometimes referred to herein as “TARP-analogs”) whose amino acid sequences are at least 90% identical to TARP (although they may have 91%, 92%, 93%, 94%, 95%, or even higher sequence identity to TARP) and which activate T-lymphocytes to cells which express TARP. Such proteins are useful as immunogens to break tolerance to PS-TCR $\gamma$  proteins.

In another embodiment, the polypeptide comprises an epitope that binds an MHC molecule, *e.g.*, an HLA molecule or a DR molecule. These molecules bind polypeptides having the correct anchor amino acids separated by about eight or nine amino acids. These peptides can be identified by inspection of the amino acid sequence of TARP and by knowledge of the MHC binding motifs, well known in the art.

TARP, immunogenic fragments thereof, and TARP analogs, can be synthesized recombinantly. Immunogenic fragments of TARP and the 58-residue TARP itself, can also be chemically synthesized by standard methods. If desired, polypeptides can also be chemically synthesized by emerging technologies. One such process is described in W. Lu *et al.*, *Federation of European Biochemical Societies Letters*. 429:31-35 (1998).



#### IV. TARP NUCLEIC ACIDS

In one aspect this invention provides an isolated, recombinant nucleic acid molecule comprising a nucleotide sequence encoding the TARP polypeptide (see, e.g., Figure 14). This nucleic acid is useful for expressing TARP, which can then be used, for example, to raise antibodies for diagnostic purposes. As noted, the nucleic acid molecule has three reading frames, each of which encodes different polypeptides defined by different open reading frames. In the embodiments contemplated herein, the reading frame of interest is the one which encodes TARP.

As noted, two reading frames were translated in *in vitro* translation systems. A nucleotide sequence of the ~1.1 kb PS-TCR $\gamma$  transcript (SEQ ID NO:13) as obtained from LNCaP cDNA and the deduced amino acid sequence when the transcript is translated from the initiation codon at nucleotide position 74 (PS-TCR $\gamma$ -1, SEQ ID NO:14) and nucleotide position 247 (PS-TCR $\gamma$ -2, SEQ ID NO:15) are presented in Fig. 1. The startpoint of transcription (underlined) is within the 10 first nucleotides of the J $\gamma$ 1.2 segment. The sequence data is available from EMBL/GenBank/DDBJ under accession number AF151103. It should be noted that it has now been determined that the actual "+1" site is the sixth nucleotide in the sequence set forth in Figure 1.

The practitioner can use this sequence to prepare PCR primers for isolating nucleotide sequences of this invention. LNCaP cells are useful sources of cDNA for sequences of the ~1.1 kb transcript. Genomic DNA from a human cell that has not undergone TCR $\gamma$  gene rearrangement, for example, cells other than T-lymphocyte precursors, are useful for longer sequences that can be processed, upon transcription, into the ~1.1 kb transcript. The sequence can be modified to engineer a nucleic acid encoding related molecules of this invention using well known techniques.

A nucleic acid comprising sequences of this invention can be cloned or amplified by *in vitro* methods, such as the polymerase chain reaction (PCR), the ligase chain reaction (LCR), the transcription-based amplification system (TAS), the self-sustained sequence replication system (3SR) and the Q $\beta$  replicase amplification system (QB). For example, a polynucleotide encoding the protein can be isolated by polymerase chain reaction of cDNA using primers based on the DNA sequence of the molecule.

A wide variety of cloning and *in vitro* amplification methodologies are well-known to persons skilled in the art. PCR methods are described in, for example, U.S. Pat. No. 4,683,195; Mullis *et al.* (1987) *Cold Spring Harbor Symp. Quant. Biol.*

51:263; and Erlich, ed., PCR TECHNOLOGY, (Stockton Press, NY, 1989). Polynucleotides also can be isolated by screening genomic or cDNA libraries with probes selected from the sequences of the desired polynucleotide under stringent hybridization conditions.

Engineered versions of the nucleic acids can be made by site-specific  
5 mutagenesis of other polynucleotides encoding the proteins, or by random mutagenesis caused by increasing the error rate of PCR of the original polynucleotide with 0.1 mM MnCl<sub>2</sub> and unbalanced nucleotide concentrations.

### 1. Expression vectors

This invention also provides expression vectors for expressing  
10 polypeptides encoded by TARP transcript. Expression vectors can be adapted for function in prokaryotes or eukaryotes by inclusion of appropriate promoters, replication sequences, markers, *etc.* for transcription and translation of mRNA. The construction of expression vectors and the expression of genes in transfected cells involves the use of molecular cloning techniques also well known in the art. Sambrook *et al.*, MOLECULAR  
15 CLONING -- A LABORATORY MANUAL, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (1989) and CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, F.M. Ausubel *et al.*, eds., (Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc.) Useful promoters for such purposes include a metallothionein promoter, a constitutive adenovirus major late promoter, a  
20 dexamethasone-inducible MMTV promoter, a SV40 promoter, a MRP polIII promoter, a constitutive MPSV promoter, a tetracycline-inducible CMV promoter (such as the human immediate-early CMV promoter), and a constitutive CMV promoter. A plasmid useful for gene therapy can comprise other functional elements, such as selectable markers, identification regions, and other genes.

25 Expression vectors useful in this invention depend on their intended use. Such expression vectors must, of course, contain expression and replication signals compatible with the host cell. Expression vectors useful for expressing bioactive conjugates include viral vectors such as retroviruses, adenoviruses and adeno-associated viruses, plasmid vectors, cosmids, and the like. Viral and plasmid vectors are preferred  
30 for transfecting mammalian cells. The expression vector pcDNA1 (Invitrogen, San Diego, CA), in which the expression control sequence comprises the CMV promoter, provides good rates of transfection and expression. Adeno-associated viral vectors are useful in the gene therapy methods of this invention.

A variety of means are available for delivering polynucleotides to cells including, for example, direct uptake of the molecule by a cell from solution, facilitated uptake through lipofection (*e.g.*, liposomes or immunoliposomes), particle-mediated transfection, and intracellular expression from an expression cassette having an expression control sequence operably linked to a nucleotide sequence that encodes the inhibitory polynucleotide. See also U.S. Patent 5,272,065 (Inouye *et al.*); METHODS IN ENZYMOLOGY, vol. 185, Academic Press, Inc., San Diego, CA (D.V. Goeddel, ed.) (1990) or M. Krieger, GENE TRANSFER AND EXPRESSION -- A LABORATORY MANUAL, Stockton Press, New York, NY, (1990). Recombinant DNA expression plasmids can also be used to prepare the polynucleotides of the invention for delivery by means other than by gene therapy, although it may be more economical to make short oligonucleotides by *in vitro* chemical synthesis.

The construct can also contain a tag to simplify isolation of the protein. For example, a polyhistidine tag of, *e.g.*, six histidine residues, can be incorporated at the amino terminal end of the protein. The polyhistidine tag allows convenient isolation of the protein in a single step by nickel-chelate chromatography.

## 2. Recombinant cells

The invention also provides recombinant cells comprising an expression vector for expression of the nucleotide sequences of this invention. Host cells can be selected for high levels of expression in order to purify the protein. The cells can be prokaryotic cells, such as *E. coli*, or eukaryotic cells. Useful eukaryotic cells include yeast and mammalian cells. The cell can be, *e.g.*, a recombinant cell in culture or a cell *in vivo*.

Cells expressing TARP are useful for active or passive immunization of subjects against cells expressing these peptides. In certain embodiments, the cells are bacterial cells. In one version of active immunization, recombinant cells are autologous cells of the subject that can present the polypeptides in association with HLA molecules. For example, antigen presenting cells are useful for this purpose. In this case, it is preferable to use "autologous cells," that is, cells derived from the subject. Such cells are MHC compatible. The TARP-encoding nucleotide sequence should be placed under the control of a constitutive promoter in such cells because one goal is to express the polypeptides in high density on the cell surface, preferably more densely than they are expressed in healthy prostate epithelial cells.

## V. METHOD OF ELICITING A CELL-MEDIATED IMMUNE RESPONSE AGAINST CELLS EXPRESSING TARP

TARP is expressed by prostate cancer cells of epithelial origin and by cells of many breast cancers. Therefore, TARP can be used as a target of intervention in the treatment of prostate cancer and TARP-expressing breast cancers, as well as a marker for cancer cells that have metastasized from the prostate or breast, respectively. This invention provides methods of treating prostate cancer and TARP-expressing breast cancers with immunotherapy. The methods involve immunizing a subject against TARP, thereby eliciting a cell-mediated immune response against cells expressing TARP.

Immunization can be active or passive. In active immunization, the immune response is elicited in the subject *in vivo*. In passive immunization, T<sub>C</sub> cells activated against the polypeptide are cultured *in vitro* and administered to the subject. Such methods may be expected to result in the destruction of healthy epithelial prostate tissue that express TARP. However, the prostate is not an essential organ. Its loss must be counterbalanced against the chance for loss of the subject's life from the prostate cancer, and the prostate may, indeed, be surgically removed prior to institution of TARP immunotherapy. Since normal breast tissue has not been found to express TARP in significant amounts, it does not appear that immunization against TARP-expressing cells will result in the loss of normal cells in women. Thus, TARP compositions may be administered to women prophylactically to provide an immune defense in the event that a TARP-expressing breast cancer develops later.

The immunizing agent can be of full-length TARP, a peptide comprising an antigenic determinant of TARP, *e.g.*, an immunogenic fragment of TARP, or a protein or peptide that is substantially identical to TARP. When one is attempting to elicit a cell-mediated immune response against TARP, preferred peptides comprising antigenic determinants are those peptides bearing a binding motif for an HLA molecule of the subject. These motifs are well known in the art. For example, HLA-A2 is a common allele in the human population. The binding motif for this molecule includes polypeptides with 9 or 10 amino acids having leucine or methionine in the second position and valine or leucine in the last positions. Based on the polypeptide sequence of TARP, one can identify amino acid sequences bearing motifs for any particular HLA molecule. Peptides comprising these motifs can be prepared by any of the typical methods (*e.g.*, recombinantly, chemically, *etc.*). Because TARP is a self protein, the preferred amino acid sequences bearing HLA binding motifs are those that encode

subdominant or cryptic epitopes. Those epitopes can be identified by a lower comparative binding affinity for the HLA molecule with respect to other epitopes in the molecule or compared with other molecules that bind to the HLA molecule.

Polypeptides that comprise an amino acid sequence from TARP that, in turn, comprise an HLA binding motif also are useful for eliciting an immune response. This is because, in part, such proteins will be processed by the cell into a peptide that can bind to the HLA molecule and that have a TARP epitope.

A complex of an HLA molecule and a peptidic antigen acts as the ligand recognized by HLA-restricted T cells (Buus, S. et al., *Cell* 47:1071, 1986; Babbitt, B. P. et al., *Nature* 317:359, 1985; Townsend, A. and Bodmer, H., *Annu. Rev. Immunol.* 7:601, 1989; Germain, R. N., *Annu. Rev. Immunol.* 11:403, 1993). Through the study of single amino acid substituted antigen analogs and the sequencing of endogenously bound, naturally processed peptides, critical residues that correspond to motifs required for specific binding to HLA antigen molecules have been identified (see, e.g., Southwood, et al., *J. Immunol.* 160:3363, 1998; Rammensee, et al., *Immunogenetics* 41:178, 1995; Rammensee et al., Sette, A. and Sidney, J. *Curr. Opin. Immunol.* 10:478, 1998; Engelhard, V. H., *Curr. Opin. Immunol.* 6:13, 1994; Sette, A. and Grey, H. M., *Curr. Opin. Immunol.* 4:79, 1992).

Furthermore, x-ray crystallographic analysis of HLA-peptide complexes has revealed pockets within the peptide binding cleft of HLA molecules which accommodate, in an allele-specific mode, residues borne by peptide ligands; these residues in turn determine the HLA binding capacity of the peptides in which they are present. (See, e.g., Madden, D.R. *Annu. Rev. Immunol.* 13:587, 1995; Smith, et al., *Immunity* 4:203, 1996; Fremont et al., *Immunity* 8:305, 1998; Stern et al., *Structure* 2:245, 1994; Jones, E.Y. *Curr. Opin. Immunol.* 9:75, 1997; Brown, J. H. et al., *Nature* 364:33, 1993.)

Accordingly, the definition of class I and class II allele-specific HLA binding motifs, or class I or class II supermotifs allows identification of regions within TARP that have the potential of binding particular HLA molecules.

Molecules with high levels of sequence identity to TARP are also useful to elicit an immune response. Such molecules can be recognized as "foreign" to the immune system, yet generate antibodies or CTLs that cross react with TARP. Molecules that have high sequence identity to TARP include non-human TCR $\gamma$  homologs, especially those from primates. TARP analogs whose amino acid sequences are at least 90%

identical to TARP (although they may have 91%, 92%, 93%, 94%, 95%, or even higher sequence identity to TARP) and which are specifically bound by antibodies which specifically bind to TARP may be used. Further useful in this regard are TARP analogs, that is, peptides whose amino acid sequences are at least 90% identical to TARP

5 (although they may have 91%, 92%, 93%, 94%, 95%, or even higher sequence identity to TARP) and which activate T-lymphocytes to cells which express TARP.

Another molecule that is substantially homologous to a TARP antigenic determinant can be made by modifying the sequence of a natural TARP epitope so that it binds with greater affinity for the HLA molecule.

10 One method of identifying genes encoding antigenic determinants is as follows: TILs from a subject with metastatic cancer are grown and tested for the ability to recognize the autologous cancer *in vitro*. These TILs are administered to the subject to identify the ones that result in tumor regression. The TILs are used to screen expression libraries for genes that express epitopes recognized by the TILs. Subjects then are  
15 immunized with these genes. Alternatively, lymphocytes are sensitized *in vitro* against antigens encoded by these genes. Then the sensitized lymphocytes are adoptively transferred into subjects and tested for their ability to cause tumor regression. Rosenberg, et al., (1997) *Immunol. Today* 1997 18:175.

The application of these molecules is now described. These methods are  
20 also described in Rosenberg *et al.* (1997) *Immunol. Today* 18:175 and Restifo *et al.* (1999) *Oncology* 11:50.

One method of invoking an immune response involves immunizing the subject with a polypeptide comprising an antigenic determinant from TARP, either alone or, more preferably, combined with an adjuvant, such as Freund's incomplete adjuvant,  
25 lipids or liposomes, gp96, Hsp70 or Hsp90. The polypeptide can be TARP, an antigenic fragment of TARP, a fusion protein comprising the antigenic determinant, or a peptide comprising a sequence substantially identical to such an antigenic determinant.

Another method involves pulsing a polypeptide comprising an epitope from TARP onto antigen presenting cells and administering the cells to the subject.

30 In another method, a recombinant virus containing a nucleic acid sequence encoding a polypeptide comprising an antigenic determinant from TARP in an expression cassette is administered to the subject. The virus optionally also can encode cytokines (*e.g.*, IL-2), a costimulatory molecule or other genes that enhance the immune response. The virus can be, for example, adenovirus, fowlpox virus or vaccinia virus. Upon

infection, the infected cells will express the TARP peptide and express the antigenic determinant on the cell surface in combination with the HLA molecule which binds peptides having the same motif as the antigenic determinant. These cells will then stimulate the activation of CTLs that recognize the presented antigen, resulting in destruction of cancer cells that also bear the determinant.

In another method, the subject is immunized with naked DNA encoding a polypeptide comprising an antigenic determinant from TARP by, *e.g.*, intramuscular, biolistic injection or linked to lipids. Such methods have been shown to result in the stimulation of a cell-mediated response against cells that express the encoded polypeptide.

In another method, a recombinant bacteria that expresses the epitope, such as *Bacillus Calmette-Guerin* (BCG), *Salmonella* or *Listeria*, optionally also encoding cytokines, costimulatory molecules or other genes to enhance the immune response, is administered to the subject.

In another method, cells expressing the antigen are administered to the subject. This includes, for example, dendritic cells pulsed with TARP epitopes, cells transfected with polypeptides comprising TARP antigenic determinants, HLA and B7 genes. The multiple transfection results in the production of several components necessary for presenting the antigenic determinant on the cell surface. In one embodiment, the molecule is a fusion protein in which the polypeptide bearing the antigenic determinant is fused to an HLA molecule (usually through a linker) so as to improve binding of the peptide to the HLA molecule. In one embodiment, the cell is an antigen presenting cell. Preferably, the cells are eukaryotic cells, more preferably, mammalian cells, more preferably, human cells, more preferably autologous human cells derived from the subject.

In another method, antigen presenting cells (APCs) are pulsed or co-incubated with peptides comprising an epitope from TARP *in vitro*. These cells are used to sensitize CD8 cells, such as tumor infiltrating lymphocytes from prostate cancer tumors or peripheral blood lymphocytes. The TILs or PBLs preferably are from the subject. However, they should at least be MHC Class-I restricted to the HLA types the subject possesses. The sensitized cells are then administered to the subject.

In a supplemental method, any of these immunotherapies is augmented by administering a cytokine, such as IL-2, IL-3, IL-6, IL-10, IL-12, IL-15, GM-CSF, interferons.

In addition to the methods for evaluating immunogenicity of peptides set forth above, immunogenicity can also be evaluated by: evaluation of primary T cell cultures from normal individuals (see, e.g., Wentworth, P. A. et al., *Mol. Immunol.* 32:603, 1995; Celis, E. et al., *Proc. Natl. Acad. Sci. USA* 91:2105, 1994; Tsai, V. et al., *J. Immunol.* 158:1796, 1997; Kawashima, I. et al., *Human Immunol.* 59:1, 1998); by immunization of HLA transgenic mice (see, e.g., Wentworth, P. A. et al., *J. Immunol.* 26:97, 1996; Wentworth, P. A. et al., *Int. Immunol.* 8:651, 1996; Alexander, J. et al., *J. Immunol.* 159:4753, 1997), and by demonstration of recall T cell responses from patients who have been effectively vaccinated or who have a tumor; (see, e.g., Rehmann, B. et al., *J. Exp. Med.* 181:1047, 1995; Doolan, D. L. et al., *Immunity* 7:97, 1997; Bertoni, R. et al., *J. Clin. Invest.* 100:503, 1997; Threlkeld, S. C. et al., *J. Immunol.* 159:1648, 1997; Diepolder, H. M. et al., *J. Virol.* 71:6011, 1997).

In choosing CTL-inducing peptides of interest for vaccine compositions, peptides with higher binding affinity for class I HLA molecules are generally preferable. Peptide binding is assessed by testing the ability of a candidate peptide to bind to a purified HLA molecule in vitro.

To ensure that a TARP analog when used as a vaccine, actually elicits a CTL response to TARP in vivo (or, in the case of class II epitopes, elicits helper T cells that cross-react with the wild type peptides), the TARP analog may be used to immunize T cells in vitro from individuals of the appropriate HLA allele. Thereafter, the immunized cells' capacity to induce lysis of TARP sensitized target cells is evaluated.

More generally, peptides from TARP or an analog thereof (a "peptide of the invention") can be synthesized and tested for their ability to bind to HLA proteins and to activate HTL or CTL responses, or both.

Conventional assays utilized to detect T cell responses include proliferation assays, lymphokine secretion assays, direct cytotoxicity assays, and limiting dilution assays. For example, antigen-presenting cells that have been incubated with a peptide can be assayed for the ability to induce CTL responses in responder cell populations.

PBMCs may be used as the responder cell source of CTL precursors. The appropriate antigen-presenting cells are incubated with peptide, after which the peptide-loaded antigen-presenting cells are then incubated with the responder cell population under optimized culture conditions. Positive CTL activation can be determined by assaying the culture for the presence of CTLs that kill radio-labeled target cells, both



specific peptide-pulsed targets as well as target cells expressing endogenously processed forms of the antigen from which the peptide sequence was derived.

A method which allows direct quantification of antigen-specific T cells is staining with Fluorescein-labelled HLA tetrameric complexes (Altman et al., Proc. Natl. Acad. Sci. USA 90:10330 (1993); Altman et al., Science 274:94 (1996)). Alternatively, staining for intracellular lymphokines, interferon- $\gamma$  release assays or ELISPOT assays, can be used to evaluate T-cell responses.

HTL activation may be assessed using such techniques known to those in the art such as T cell proliferation and secretion of lymphokines, e.g. IL-2 (see, e.g. Alexander et al., Immunity 1:751-761 (1994)).

## VI. ANTIBODIES AGAINST TARP

In one aspect this invention provides a composition comprising an antibody that specifically binds TARP. Antibodies preferably have affinity of at least  $10^6 M^{-1}$ ,  $10^7 M^{-1}$ ,  $10^8 M^{-1}$ , or  $10^9 M^{-1}$ . This invention contemplates both polyclonal and monoclonal antibody compositions.

A number of immunogens can be used to produce antibodies that specifically bind TARP. Full-length TARP is a suitable immunogen. Typically, the immunogen of interest is a peptide of at least about 3 amino acids, more typically the peptide is at least 5 amino acids in length, preferably, the fragment is at least 10 amino acids in length and more preferably the fragment is at least 15 amino acids in length. The peptides can be coupled to a carrier protein (e.g., as a fusion protein), or are recombinantly expressed in an immunization vector. Antigenic determinants on peptides to which antibodies bind are typically 3 to 10 amino acids in length. Naturally occurring polypeptides are also used either in pure or impure form.

Recombinant polypeptides are expressed in eukaryotic or prokaryotic cells and purified using standard techniques. The polypeptide, or a synthetic version thereof, is then injected into an animal capable of producing antibodies. Either monoclonal or polyclonal antibodies can be generated for subsequent use in immunoassays to measure the presence and quantity of the polypeptide.

Methods for producing polyclonal antibodies are known to those of skill in the art. In brief, an immunogen, preferably a purified polypeptide, a polypeptide coupled to an appropriate carrier (e.g., GST, keyhole limpet hemocyanin, *etc.*), or a polypeptide incorporated into an immunization vector such as a recombinant vaccinia virus (see, U.S.

Patent No. 4,722,848) is mixed with an adjuvant and animals are immunized with the mixture. The animal's immune response to the immunogen preparation is monitored by taking test bleeds and determining the titer of reactivity to the polypeptide of interest. When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera are prepared. Further fractionation of the antisera to enrich for antibodies reactive to the polypeptide is performed where desired. See, *e.g.*, Coligan (1991) *CURRENT PROTOCOLS IN IMMUNOLOGY* Wiley/Greene, NY; and Harlow and Lane (1989) *ANTIBODIES: A LABORATORY MANUAL* Cold Spring Harbor Press, NY.

Antibodies, including binding fragments and single chain recombinant versions thereof, against predetermined fragments of TARP are raised by immunizing animals, *e.g.*, with conjugates of the fragments with carrier proteins as described above.

Monoclonal antibodies are prepared from cells secreting the desired antibody. These antibodies are screened for binding to normal or modified polypeptides, or screened for agonistic or antagonistic activity. In some instances, it is desirable to prepare monoclonal antibodies from various mammalian hosts, such as mice, rodents, primates, humans, *etc.* Description of techniques for preparing such monoclonal antibodies are found in, *e.g.*, Stites *et al.* (eds.) *BASIC AND CLINICAL IMMUNOLOGY* (4th ed.) Lange Medical Publications, Los Altos, CA, and references cited therein; Harlow and Lane, *Supra*; Goding (1986) *Monoclonal Antibodies: Principles and Practice* (2d ed.) Academic Press, New York, NY; and Kohler and Milstein (1975) *Nature* 256: 495-497.

Other suitable techniques involve selection of libraries of recombinant antibodies in phage or similar vectors. See, Huse *et al.* (1989) *Science* 246: 1275-1281; and Ward, *et al.* (1989) *Nature* 341: 544-546.

Also, recombinant immunoglobulins may be produced. See, , U.S. Patent No. 4,816,567 (Cabilly); and Queen *et al.* (1989) *Proc. Nat'l Acad. Sci. USA* 86: 10029-10033.

Frequently, the polypeptides and antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Thus, an antibody used for detecting an analyte can be directly labeled with a detectable moiety, or may be indirectly labeled by, for example, binding to the antibody a secondary antibody that is, itself directly or indirectly labeled.

The antibodies of this invention are also used for affinity chromatography in isolating TARP. Columns are prepared, *e.g.*, with the antibodies linked to a solid support, *e.g.*, particles, such as agarose, Sephadex, or the like, where a cell lysate is passed through the column, washed, and treated with increasing concentrations of a mild denaturant, whereby purified TARP is released.

An alternative approach is the generation of humanized immunoglobulins by linking the CDR regions of non-human antibodies to human constant regions by recombinant DNA techniques. See United States patent 5,585,089 (Queen *et al.*).

A further approach for isolating DNA sequences which encode a human monoclonal antibody or a binding fragment thereof is by screening a DNA library from human B cells according to the general protocol outlined by Huse *et al.*, *Science* 246:1275-1281 (1989) and then cloning and amplifying the sequences which encode the antibody (or binding fragment) of the desired specificity. The protocol described by Huse is rendered more efficient in combination with phage display technology. See, *e.g.*, WO 91/17271 (Dower *et al.*) and WO 92/01047 (McCafferty *et al.*). Phage display technology can also be used to mutagenize CDR regions of antibodies previously shown to have affinity for TARP. Antibodies having improved binding affinity are selected.

In another embodiment of the invention, fragments of antibodies against TARP or protein analogs are provided. Typically, these fragments exhibit specific binding to TARP similar to that of a complete immunoglobulin. Antibody fragments include separate heavy chains, light chains Fab, Fab' F(ab')<sub>2</sub> and Fv. Fragments are produced by recombinant DNA techniques, or by enzymic or chemical separation of intact immunoglobulins.

## VII. CHIMERIC MOLECULES THAT TARGET TARP

This invention provides chimeric molecules that target TARP. The chimeric molecules comprise a targeting moiety and an effector moiety. The chimeric proteins are useful in the detection of the polypeptide and cells that bear it.

### A. Targeting Moiety

The chimeric molecules of this invention comprise a targeting moiety. The targeting moiety comprises a ligand that specifically binds to TARP. Preferred ligands are antibodies, as that term is used here, including binding fragments of antibodies. However, other natural ligands for these molecules also can be used.

### B. Effector moiety

The effector moiety may be another specific binding moiety such as an antibody, a growth factor, or a ligand. The chimeric molecule will then act as a highly specific bifunctional linker. This linker may act to bind and enhance the interaction  
5 between cells or cellular components to which the fusion protein binds.

In still yet another embodiment the effector molecule may be a pharmacological agent (*e.g.* a drug) or a vehicle containing a pharmacological agent. Thus, the moiety that specifically binds to TARP may be conjugated to a drug such as vinblastine, doxorubicin, genistein (a tyrosine kinase inhibitor), an antisense molecule,  
10 and other pharmacological agents known to those of skill in the art, thereby specifically targeting the pharmacological agent to tumor cells.

Alternatively, the targeting molecule may be bound to a vehicle containing the therapeutic composition. Such vehicles include, but are not limited to liposomes, micelles, various synthetic beads, and the like.

15 One of skill in the art will appreciate that the chimeric molecules of the present invention may include multiple targeting moieties bound to a single effector or conversely, multiple effector molecules bound to a single targeting moiety. In still other embodiments, the chimeric molecules may include both multiple targeting moieties and multiple effector molecules. Detectable labels suitable for use as the effector molecule  
20 component of the chimeric molecules of this invention include any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means all as described above.

One of skill will appreciate that the targeting molecule and effector molecules may be joined together in any order. Thus, where the targeting molecule is a  
25 polypeptide, the effector molecule may be joined to either the amino or carboxy termini of the targeting molecule. The targeting molecule may also be joined to an internal region of the effector molecule, or conversely, the effector molecule may be joined to an internal location of the targeting molecule, as long as the attachment does not interfere with the respective activities of the molecules.

30 The targeting molecule and the effector molecule may be attached by any of a number of means well known to those of skill in the art. Typically the effector molecule is conjugated, either directly or through a linker (spacer), to the targeting molecule. However, where both the effector molecule and the targeting molecule are

polypeptides it is preferable to recombinantly express the chimeric molecule as a single-chain fusion protein.

In one embodiment, the targeting molecule is chemically conjugated to the effector molecule (*e.g.* a cytotoxin, a label, a ligand, or a drug or liposome). Means of chemically conjugating molecules are well known to those of skill. The procedure for attaching an agent to an antibody or other polypeptide targeting molecule will vary according to the chemical structure of the agent. Polypeptides typically contain variety of functional groups; *e.g.*, carboxylic acid (COOH) or free amine (-NH<sub>2</sub>) groups, which are available for reaction with a suitable functional group on an effector molecule to bind the effector thereto. Alternatively, the targeting molecule and/or effector molecule may be derivatized to expose or attach additional reactive functional groups. The derivitization may involve attachment of any of a number of linker molecules such as those available from Pierce Chemical Company, Rockford Illinois.

A bifunctional linker having one functional group reactive with a group on a particular agent, and another group reactive with an antibody, may be used to form the desired immunoconjugate. Alternatively, derivitization may involve chemical treatment of the targeting molecule, *e.g.*, glycol cleavage of the sugar moiety of a the glycoprotein antibody with periodate to generate free aldehyde groups. The free aldehyde groups on the antibody may be reacted with free amine or hydrazine groups on an agent to bind the agent thereto. (See U.S. Patent No. 4,671,958). Procedures for generation of free sulfhydryl groups on polypeptide, such as antibodies or antibody fragments, are also known (See U.S. Pat. No. 4,659,839).

Many procedures and linker molecules for attachment of various compounds including radionuclide metal chelates, toxins and drugs to proteins such as antibodies are known. See, for example, European Patent Application No. 188,256; U.S. Patent Nos. 4,671,958, 4,659,839, 4,414,148, 4,699,784; 4,680,338; 4,569,789; and 4,589,071; and Borlinghaus *et al.* *Cancer Res.* 47: 4071-4075 (1987). In particular, production of various immunotoxins is well-known within the art and can be found, for example in "Monoclonal Antibody-Toxin Conjugates: Aiming the Magic Bullet," Thorpe *et al.*, MONOCLONAL ANTIBODIES IN CLINICAL MEDICINE, Academic Press, pp. 168-190 (1982), Waldmann, *Science*, 252: 1657 (1991), U.S. Patent Nos. 4,545,985 and 4,894,443.

Where the targeting molecule and/or the effector molecule is relatively short (*i.e.*, less than about 50 amino acids) they may be synthesized using standard

chemical peptide synthesis techniques. Where both molecules are relatively short the chimeric molecule may be synthesized as a single contiguous polypeptide. Alternatively the targeting molecule and the effector molecule may be synthesized separately and then fused by condensation of the amino terminus of one molecule with the carboxyl terminus of the other molecule thereby forming a peptide bond. Alternatively, the targeting and effector molecules may each be condensed with one end of a peptide spacer molecule thereby forming a contiguous fusion protein.

Solid phase synthesis in which the C-terminal amino acid of the sequence is attached to an insoluble support followed by sequential addition of the remaining amino acids in the sequence is the preferred method for the chemical synthesis of the polypeptides of this invention. Techniques for solid phase synthesis are described by Barany and Merrifield, *Solid-Phase Peptide Synthesis*; pp. 3-284 in *The Peptides: Analysis, Synthesis, Biology*. Vol. 2: *Special Methods in Peptide Synthesis, Part A.*, Merrifield, *et al. J. Am. Chem. Soc.*, 85: 2149-2156 (1963), and Stewart *et al.*, *Solid Phase Peptide Synthesis*, 2nd ed. Pierce Chem. Co., Rockford, Ill. (1984).

In a preferred embodiment, the chimeric fusion proteins are synthesized using recombinant DNA methodology. Generally this involves creating a DNA sequence that encodes the fusion protein, placing the DNA in an expression cassette under the control of a particular promoter, expressing the protein in a host, isolating the expressed protein and, if required, renaturing the protein.

DNA encoding the fusion proteins of this invention may be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences or direct chemical synthesis by methods such as the phosphotriester method of Narang *et al. Meth. Enzymol.* 68: 90-99 (1979); the phosphodiester method of Brown *et al.*, *Meth. Enzymol.* 68: 109-151 (1979); the diethylphosphoramidite method of Beaucage *et al.*, *Tetra. Lett.*, 22: 1859-1862 (1981); and the solid support method of U.S. Patent No. 4,458,066.

While the two molecules are preferably essentially directly joined together, one of skill will appreciate that the molecules may be separated by a peptide spacer consisting of one or more amino acids. Generally the spacer will have no specific biological activity other than to join the proteins or to preserve some minimum distance or other spatial relationship between them. However, the constituent amino acids of the spacer may be selected to influence some property of the molecule such as the folding, net charge, or hydrophobicity.

The nucleic acid sequences encoding the fusion proteins may be expressed in a variety of host cells, including *E. coli*, other bacterial hosts, yeast, and various higher eukaryotic cells such as the COS, CHO and HeLa cells lines and myeloma cell lines. The recombinant protein gene will be operably linked to appropriate expression control sequences for each host. For *E. coli* this includes a promoter such as the T7, trp, or lambda promoters, a ribosome binding site and preferably a transcription termination signal. For eukaryotic cells, the control sequences will include a promoter and preferably an enhancer derived from immunoglobulin genes, SV40, cytomegalovirus, *etc.*, and a polyadenylation sequence, and may include splice donor and acceptor sequences. The plasmids and vectors of the invention can be transferred into the chosen host cell by well-known methods such as calcium chloride transformation for *E. coli* and calcium phosphate treatment or electroporation for mammalian cells.

Once expressed, the recombinant fusion proteins can be purified according to standard procedures of the art, including ammonium sulfate precipitation, affinity columns, column chromatography, gel electrophoresis and the like (see, generally, R. Scopes, *PROTEIN PURIFICATION*, Springer--Verlag, N.Y. (1982), Deutscher, *METHODS IN ENZYMOLOGY* Vol. 182: *Guide to Protein Purification.*, Academic Press, Inc. N.Y. (1990)). Substantially pure compositions of at least about 90 to 95% homogeneity are preferred, and 98 to 99% or more homogeneity are most preferred for pharmaceutical uses. Once purified, partially or to homogeneity as desired, the polypeptides may then be used therapeutically.

## VIII. METHODS OF DETECTING CELLS THAT EXPRESS TARP

In another aspect, this invention provides methods of detecting cells that express TARP. The methods involve detecting either a TARP transcript or polypeptide. Because prostate cancer cells of epithelial origin and many breast cancer cells express TARP, methods of detection are useful in the detection of prostate cancer and of TARP-expressing breast cancers. In particular, prostate cancer cells and many breast cancer cells can be distinguished from other cells by the expression of TARP.

Tissue samples can be selected from any likely site of primary or metastatic cancer including the prostate or the breast, respectively, and distal sites such as the lymph nodes and other organs. Persons of skill in the art are aware that men, as well as women, suffer from breast cancer. Breast cancer in men is relatively rare, representing only about 1% of all breast cancer cases. Because it is uncommon, however, it is

frequently diagnosed at a later stage, which affects the chances for survival. Accordingly, improved diagnosis of breast cancer in men is desirable.

In one method, a biopsy is performed on the subject and the collected tissue is tested *in vitro*. Typically, the cells are disrupted by lysing, sonic disruption, osmotic pressure, freezing and thawing, enzymatic treatment, or other means routine in the art to render the proteins of the nucleus accessible without denaturing them. The cellular contents (or the nuclear contents, if the contents have been fractionated) are then contacted, for example, with an anti-TARP antibody. Any immune complexes which result indicate the presence of TARP in the biopsied sample. To facilitate such detection, the antibody can be radiolabeled or coupled to an effector molecule which is radiolabelled. In another method, the cells can be detected *in vivo* using typical imaging systems. For example, the method can involve the administration to a subject of a labeled composition capable of reaching the cell nucleus. Then, the localization of the label is determined by any of the known methods for detecting the label. Any conventional method for visualizing diagnostic imaging can be used. For example, paramagnetic isotopes can be used for MRI.

### Detection of TARP

TARP can be identified by any methods known in the art. In one embodiment, the methods involve detecting the polypeptide with a ligand that specifically recognizes the polypeptide (*e.g.*, an immunoassay). The antibodies of the invention are particularly useful for specific detection of TARP. A variety of antibody-based detection methods are known in the art. These include, for example, radioimmunoassay, sandwich immunoassays (including ELISA), immunofluorescent assays, Western blot, affinity chromatography (affinity ligand bound to a solid phase), and *in situ* detection with labeled antibodies. Another method for detecting TARP involves identifying the polypeptide according to its mass through, for example, gel electrophoresis, mass spectrometry or HPLC. Subject samples can be taken from any number of appropriate sources, such as saliva, peritoneal fluid, blood or a blood product (*e.g.*, serum), urine, tissue biopsy (*e.g.*, lymph node tissue), *etc.*

TARP can be detected in cells *in vitro*, in samples from biopsy and *in vivo* using imaging systems described above.



### Detection of transcript encoding TARP

Cells that express TARP transcript can be detected by contacting the sample with a nucleic acid probe that specifically hybridizes with the transcript, and detecting hybridization. This includes, for example, methods of *in situ* hybridization, in which a labeled probe is contacted with the sample and hybridization is detected by detecting the attached label. However, the amounts of transcript present in the sample can be small. Therefore, other methods employ amplification, such as RT-PCR. In these methods, probes are selected that function as amplification primers which specifically amplify the TARP sequences from mRNA. Then, the amplified sequences are detected using typical methods.

The probes are selected to specifically hybridize with TARP transcripts. Generally, complementary probes are used. However, probes need not be exactly complementary if they have sufficient sequence homology and length to hybridize under stringent conditions.

## IX. PHARMACEUTICAL COMPOSITIONS

In another aspect, this invention provides pharmaceutical compositions that comprise a pharmaceutically acceptable carrier and a composition of this invention.

In one embodiment, the pharmaceutical composition comprises TARP, an immunogenic fragment thereof, such as a polypeptide comprising a TARP epitope, or a TARP analog, in an amount effective to elicit a cell-mediated immune response or a humoral response in a subject, *e.g.*, a polypeptide bearing an MHC binding motif. Such pharmaceutical compositions are useful as vaccines in the therapeutic methods of this invention and for preparing antibodies.

In another embodiment, the pharmaceutical composition comprises a nucleic acid molecule comprising a nucleotide sequence encoding a TARP polypeptide in an amount effective to elicit an immune response against cells expressing TARP in a subject. Such composition also are useful in the therapeutic methods of this invention.

In another embodiment, the pharmaceutical composition comprises a ribozyme which can specifically cleave a nucleotide sequence encoding TARP, an antisense molecule which can bind to such a nucleic acid, or an expression cassette comprising a nucleic acid encoding TARP, to modulate expression of TARP in a cell of interest.

In yet another embodiment, the pharmaceutical composition may comprise a chimeric molecule comprising a targeting molecule and a detector molecule to detect cells expressing TARP. If the detector molecule is one capable of binding specifically to a nucleic acid encoding TARP (such as a DNA binding protein which can bind specifically to DNA encoding TARP), then the composition can be used to detect cells which express that nucleic acid.

The pharmaceutical compositions of this invention can be prepared in unit dosage forms for administration to a subject. The amount and timing of administration are at the discretion of the treating physician to achieve the desired purposes.

## EXAMPLES

### EXAMPLE 1. DETECTION OF T-CELL RECEPTOR $\gamma$ -CHAIN IN PROSTATE CELLS

We identified expression of T-cell receptor  $\gamma$ -chain (TCR $\gamma\delta$  mRNA in human prostate and showed that it originates from epithelial cells of the prostate and not from infiltrating T-lymphocytes. In contrast, the T-cell receptor  $\delta$ -chain (TCR $\gamma\delta$  gene is silent in human prostate. The major TCR $\gamma$  transcript in prostate has a different size than the transcript expressed in thymus, spleen and blood leukocytes. It is expressed in normal prostate epithelium, adenocarcinoma of the prostate and the prostatic adenocarcinoma cell line LNCaP. The RNA originates from an unrearranged TCR $\gamma$  locus and it is initiated within the intronic sequence directly upstream of the J $\gamma$ 1.2 gene segment. The prostate-specific TCR $\gamma$  transcript consists of the J $\gamma$ 1.2 and C $\gamma$ 1 gene segments, it has untranslated sequence including a polyadenylation signal and poly(A) sequence at the 3' end. The finding that prostate epithelial cells express a high level of a transcript from a gene that was thought to be exclusively expressed by T-lymphocytes is novel and highly unexpected.

#### 1. Materials and Methods

##### RNA dot blot and Northern blot hybridizations

RNA dot blot (RNA master blot, Clontech, Palo Alto, CA), and Northern blot (MTN, Clontech, Palo Alto, CA), were performed on a variety of human tissues.

Northern blot was also performed on mRNA from prostate adenocarcinoma cell lines, LNCaP and PC-3 (ATCC, Rockville, MD). Isolation of poly(A) RNA was carried out using the FastTrack kit (InVitrogen, Carlsbad, CA). RNA was electrophoresed on a 1% agarose gel and transferred to nylon-based membranes (GeneScreen Plus, DuPont, Wilmington, DE), according to established procedures. Ausubel, *supra*. A cDNA probe specific for the untranslated 3' end (3' UTR) of the TCR $\gamma$  transcript was made from EST plasmid ng79d11 (Genome Systems, St. Louis, MO). A probe specific for the constant domain of the TCR $\gamma$  transcript (TCR C $\gamma$ ) was made from LNCaP cDNA and a probe for the constant domain of the TCR $\delta$  transcript (TCR C $\delta$ ) was made from a TCR $\delta$  plasmid. A human  $\beta$ -actin probe was used as a quantity control of the mRNA preparations. Probes were labeled with  $^{32}\text{P}$  by random primer extension (Lofstrand Labs Limited, Gaithersburg, MD) to a specific activity of 1  $\mu\text{Ci}/\text{ng}$ . The RNA membranes were blocked for 2 hours at 45° C in hybridization solution containing 50% formamide (Hybrisol I, Oncor, Gaithersburg, MD) and then probed for 15 hours at 45° C with 20  $\mu\text{Ci}$  cDNA in 20 ml hybridization solution. The membranes were washed twice for 15 minutes at room temperature in 2xSSC/0.1%SDS and twice for 20 minutes at 55-65° C in 0.1%SSC/0.1%SDS. The membranes were exposed to an imaging film (X-OMAT, Kodak, Rochester, NY) at -80° C before development.

## 20 RNA in situ hybridization

The TCR $\gamma$  constant domain and the TCR $\gamma$  untranslated 3' end nucleotide sequence was amplified by reverse transcriptase PCR (RT-PCR) from LNCaP mRNA, cloned into pBluescript II SK (Stratagene, La Jolla, CA) and verified by DNA sequencing. Anti-sense and sense TCR $\gamma$   $^{35}\text{S}$ -riboprobes were made by T7 and T3 RNA polymerase, respectively. Paraffin blocks of 8 archived cases of prostatic transurethral resection specimens from the NCI were retrieved. Cases were selected which included both malignant and benign prostatic ducts. Average age of the cases was 69 and Gleason scores of the tumor ranged from 3+3=6/10 to 4+5=9/10. The blocks were processed on glass slides and hybridized using the riboprobes (Molecular Histology, Gaithersburg, MD). Following hybridization the slides were counterstained with Hematoxylin and Eosin and examined using a Zeiss Axiophot Microscope equipped with a variable condenser providing bright field and dark field.

### RT-PCR analysis

Single stranded cDNAs were prepared from 150-250 ng of LNCaP and PC-3 poly(A) mRNA, respectively, using oligo-dT priming (Pharmacia-Biotech, Piscataway, NJ). PCR primers were designed to amplify different portions of the TCR $\gamma$  transcript. In order to amplify cDNA only and not trace amounts of genomic DNA, which may be present in the mRNA preparations, primer pairs were always combined to generate PCR products spanning two or more exons. One PCR was set up to amplify either of the two TCR $\gamma$  constant domain genes, C $\gamma$ 1 or C $\gamma$ 2, with a forward primer in exon CI (TCRC $\gamma$ .F) and a reverse primer in exon CIII (TCRC $\gamma$ .R4), Figure 8. Variable to constant domain-spanning PCRs were set up using forward primers, specific to each of the four subgroups of TCR $\gamma$  variable gene segments (TCRV $\gamma$ I.F, TCRV $\gamma$ II.F, TCRV $\gamma$ III.F, TCRV $\gamma$ IV.F) in combination with a reverse primer in the TCR $\gamma$  constant gene segment (TCRC $\gamma$ .R1), Figure 8. Wax-mediated, hot-start PCRs were conducted for 30 cycles using high-fidelity PCR components (Expand, Boehringer-Mannheim, Indianapolis, IN). The PCR products were analyzed on 1.2% agarose gels with 0.5  $\mu$ g/ml of EtBr. Specific PCR products were gel purified (Qiagen, Valencia, CA), T/A cloned (InVitrogen, Carlsbad, CA) and sequenced on an automated capillary sequencer (Perkin Elmer Applied Systems, Foster City, CA), using Perkin-Elmer's dRhodamine terminator cycle sequencing kit.

### Analysis of TCR $\gamma$ VJ gene rearrangement

Genomic DNA was prepared from  $5 \times 10^7$  LNCaP cells according to established procedures. Ausubel, *supra*. A set of 12 PCRs was performed, each with a forward primer from one of the four subgroup of V $\gamma$  gene segments (TCRV $\gamma$ I.F, TCRV $\gamma$ II.F, TCRV $\gamma$ III.F, TCRV $\gamma$ IV.F) in combination with a reverse primer from one of the three J $\gamma$ 1 gene segments (TCRJ $\gamma$ 1.1.R, TCRJ $\gamma$ 1.2.R, TCRJ $\gamma$ 1.3.R), Figure 8. Hot-Start PCRs were conducted for 30 cycles using 500 ng of genomic DNA and the PCRs were examined on 1.2% agarose gels with 0.5  $\mu$ g/ml of EtBr. Human placenta DNA (Clontech, Palo Alto, CA) was used as a positive control of the primers and PCR amplification of J $\gamma$ 1.1 to J $\gamma$ 1.2 genomic DNA was performed as a positive control of the template.

### Primer-extension analysis of RNA

The startpoint of the prostate TCR $\gamma$  transcript was determined by primer-extension analysis of LNCaP mRNA. Five  $\mu$ g of mRNA was mixed with 0.08 pmol of

<sup>32</sup>P-end labeled TCRC $\gamma$ .R2 primer, annealing 48-75 nucleotides from the 5' end of C $\gamma$ 1. The analysis was carried out using 20 U of MMLV-reverse transcriptase (Superscript, Gibco-BRL, Gaithersburg, MD), according to established procedure. C.P. George et al/ (1996) "Primer-extension analysis of RNA" *In A LABORATORY GUIDE TO RNA,*  
5 *ISOLATION, ANALYSIS AND SYNTHESIS.* ed. P.A. Krieg. (Wiley-Liss, Inc., New York, NY), pp. 133-139. The sample was electrophoresed on a 6% polyacrylamide-urea DNA sequencing gel in parallel with a <sup>32</sup>P-end labeled molecular weight marker (*Msp*I digested pBR322, Lofstand Labs Limited, Gaithersburg, MD). After electrophoresis the gel was blotted to Whatman paper, dried and subjected to autoradiography.

10

### 5'-RACE PCR analysis

Double-stranded cDNA was made from 500 ng of LNCaP poly(A) mRNA using the Marathon cDNA amplification kit (Clontech, Palo Alto, CA) and 25 pmole of the TCR $\gamma$  gene-specific primer (TCRC $\gamma$ .R3), Figure 8. Marathon-adaptors were then  
15 ligated to the ends of the synthesized cDNA. Rapid amplification of the 5'-cDNA ends (5'-RACE) PCR was conducted using a gene-specific primer (TCRC $\gamma$ .R2), Figure 8, annealing upstream of the primer used for reverse transcription, and an adaptor-specific primer. Hot start conditions were applied (Advantage, Clontech, Palo Alto, CA) and the PCR products were analyzed and cloned as described for the RT-PCR analyses. DNA  
20 from the 5'-RACE PCR analytic gel was transferred to a nylon membrane and a <sup>32</sup>P-end labeled primer (TCRC $\gamma$ .R1), Figure 8, hybridizing further upstream was applied to identify possible bands not detected by EtBr/UV.

### *In vitro* transcription-coupled translation

25 The complete prostate TCR $\gamma$  transcript, as obtained by RT-PCR and 5'-RACE PCR, was amplified by RT-PCR, cloned into pBluescript II SK (Stratagene, La Jolla, CA), sequenced and examined in an *in vitro* transcription-coupled translation system, using T7 RNA polymerase and wheat germ extract (TNT, Promega, Madison, WI). <sup>35</sup>S-Met (ICN, Costa Mesa, CA) was incorporated in the reaction for visualization of  
30 translated products. The reaction was analyzed under reducing condition on a polyacrylamide gel (16.5% Tris/Tricine, BioRad, Hercules, CA) together with a pre-stained marker (Gibco-BRL, Gaithersburg, MD). The gel was dried and subjected to autoradiography.

## RESULTS

### A. Prostate ESTs representing TCR $\gamma$ were identified by database analysis.

We identified 23 TCR $\gamma$  ESTs, from 20 cDNA clones, derived from 6 tumor  
5 and 2 normal prostate cDNA libraries. The TCR $\gamma$  composite sequence from assembly of  
prostate ESTs has 76 nucleotides of TCR $\gamma$  constant domain sequence, 448 nucleotides of  
untranslated 3' region sequence and poly(A) sequence. By alignment of the prostate ESTs  
to mature TCR $\gamma$  transcripts from cell lines established from peripheral blood T-  
lymphocytes (GenBank Acc. No. M16768, M16804 and M30894) we found that the  
10 prostate EST composite sequence is identical to the TCR $\gamma$  transcript from peripheral  
blood T-lymphocytes. The dbEST database analysis indicates that the TCR $\gamma$  gene is  
highly transcribed in human prostate.

### B. Expression of TCR $\gamma$ (3' UTR) in human prostate verified by RNA dot blot.

15 To analyze the transcriptional activity TCR $\gamma$  gene in human prostate, a  
cDNA probe from the untranslated 3' end (3' UTR) of the TCR $\gamma$  transcript was assayed on  
mRNA from 50 different human tissues, Figure 2A. We verified that normal prostate  
(position C7) expresses TCR $\gamma$  mRNA and we further observed that prostate has by far the  
strongest expression of all tissues represented on the dot blot membrane. TCR $\gamma$  gene  
20 expression was also found in small intestine (E3), spleen (E4), thymus (E5), peripheral  
leukocyte (E6), lymph node (E7), bone marrow (E8), and lung (F2).

### C. Northern shows two size-specific TCR $\gamma$ transcripts in human prostate.

Northern blot hybridization using the 3' UTR probe revealed that prostate  
has two TCR $\gamma$  transcripts of approximately 1.1 and 2.8 kb, Figure 2B (lane 3) while the  
25 predominant transcript in spleen, thymus, small intestine and blood leukocytes is 1.5 kb.  
A transcript size of 1.5 kb is consistent with TCR $\gamma$  mRNA from  $\gamma\delta$  T-lymphocytes  
(GenBank Acc. No. M16768, M16804, (Kragel *et al.*, *Science* 237, 64-67 (1987));  
M30894, (Littman *et al.* *Nature* 326, 85-88 (1987)). Since the database analysis indicated  
that a constant domain of TCR $\gamma$  is part of the prostate transcript, we also used a TCR $\gamma$   
30 constant domain probe (TCR C $\gamma$ ). We found the same 1.1 kb and 2.8 kb bands in the  
prostate, Figure 3A (lane 3).

**D. Prostate cells expressing TCR $\gamma$  do not express TCR $\delta$  or CD3 transcripts.**

TCR  $\gamma$ -chain protein is normally co-expressed with the TCR  $\delta$ -chain protein. Since the TCR $\gamma$  gene is transcriptionally active in human prostate, we went on to  
5 analyze the transcriptional activity of the TCR $\delta$  gene. The dbEST was analyzed (<http://www.ncbi.nlm.nih.gov/BLAST>) using the TCR $\delta$  transcript nucleotide sequence. ESTs from prostate cDNA libraries did not match any part of the TCR  $\delta$ -chain transcript. Furthermore, Northern blot analysis did not detect any prostate expression of TCR $\delta$  mRNA, Figure 3B (lane 3). We conclude that the TCR $\delta$  gene is silent in prostate. As  
10 expected, TCR $\delta$  transcripts are expressed in spleen, thymus and blood leukocytes, Figure 3B.

**E. LNCaP cells, but not PC-3 cells, express the prostate-specific TCR $\gamma$  transcripts.**

Given that TCR $\gamma$  mRNA is expressed in normal prostate, we next analyzed  
15 whether it is also expressed in prostate cancer. The prostate-specific 1.1 kb transcript was found in mRNA preparations from LNCaP, but not in mRNA preparations from PC-3, Figure 3C. The prostate-specific 2.8 kb transcript, expressed in normal prostate, is also present in LNCaP although to a much lesser degree.

**A. RNA in situ hybridization shows TCR $\gamma$  expression in prostate epithelial cells.**

The prostate consists of acinar glandular tissue with variable and mixed population of simple duct lining epithelial cells, ranging to complex hyperplastic ducts in the glandular compartments. These compartments are tightly connected to smooth muscle cells, fibroblasts and other cell types in the prostate stroma. To determine the cellular  
25 localization of the human prostate TCR $\gamma$  expression, RNA in situ hybridization was carried out with TCR(C $\gamma$ -3' UTR) sense and anti-sense riboprobes. We found that TCR $\gamma$  mRNA is highly expressed in epithelial cells within the acinar ducts of the prostate while stromal cells and other cell types in the prostate are negative, Figure 4A, 4C. TCR $\gamma$  expression was also detected in hyperplastic and neoplastic areas of the prostate. The  
30 expression in benign and neoplastic acinar epithelium is comparable. TCR $\gamma$  expression could not be observed in human kidney tissue, Figure 4E, or in human brain.

**G. The prostate TCR $\gamma$  transcript contains C $\gamma$ 1 but not any VJ $\gamma$  genes.**

After we had established the TCR  $\gamma\delta$  expression profile in the prostate we went on to characterize the predominant, 1.1 kb, prostate-specific TCR $\gamma$  transcript. The LNCaP cell line was used for the characterization since one can not exclude the possibility of mRNA contamination from infiltrating T-cells in the mRNA preparations extracted from bulk prostate tissue. We knew from database analysis that the 3' end sequence of the prostate TCR $\gamma$  transcript is identical to that from peripheral blood leukocytes and that the location of the polyadenylation signal is identical. Therefore, the difference in transcript size between prostate and leukocytes is due to sequence differences upstream of the stretch identified by the prostate ESTs. An RT-PCR set up to amplify the constant domain portion of the TCR $\gamma$  transcript identified the TCRC $\gamma$ 1 gene. The slightly larger TCRC $\gamma$ 2 is not expressed in LNCaP. Variable domain (V $\gamma$ ) to constant domain (C $\gamma$ )-spanning RT-PCRs did not yield any product, indicating that V $\gamma$  is not part of the prostate-specific TCR $\gamma$  transcript.

**H. LNCaP has not undergone VJ gene rearrangement in the TCR $\gamma$  locus.**

Since RT-PCRs intending to amplify the variable domain of TCR $\gamma$  did not yield any product we next analyzed the TCR $\gamma$  locus. During the development of  $\gamma\delta$  T-cells the TCR loci undergo V(D)J gene rearrangements to bring together the gene segments that make up the variable domain of the receptor. To address whether LNCaP cells have undergone TCR $\gamma$  VJ gene rearrangement PCRs were carried out on genomic DNA using combinations of TCRV $\gamma$  and TCRJ $\gamma$  primers, to cover every possible rearrangement (see Materials and Methods). None of the primer combinations yielded any PCR product showing that LNCaP cells have not undergone VJ gene rearrangement of the TCR $\gamma$  locus. The fact that TCR $\gamma$  VJ rearrangement has not taken place in prostate epithelial cells, shows that the prostate expression is different from that of mature  $\gamma\delta$  T-lymphocytes.

**I. Prostate epithelial cells express a TCR (JC) $\gamma$  transcript.**

Since the identified prostate TCR $\gamma$  transcripts consist of C $\gamma$  but not of any V $\gamma$  gene segment, we next analyzed what sequence is upstream of C $\gamma$ 1. RNA primer-extension and 5'RACE PCR were carried out to obtain the startpoint of transcription. The primer-extension experiment conducted on LNCaP mRNA, showed a predominant band of approximately 128 nucleotides with minor bands in the 130-135 nucleotide area, Figure 5. Since the reverse transcription started 75 bases from the 5' end of C $\gamma$ 1 (see



Materials and Methods) the transcript has about 53 nucleotides upstream of C $\gamma$ 1. The 5' RACE PCR conducted on LNCaP cDNA revealed one specific PCR product. The amplified product was found to contain a J $\gamma$ 1.2 gene segment, correctly spliced to the C $\gamma$ 1 gene segments. A number of clones isolated by RACE PCR were sequenced. They initiated close to the start site defined by the primer extension experiment. A somewhat variable starting point of transcription is consistent with the identification of minor bands slightly larger than the predominant one in the primer-extension experiment. An illustration of how the prostate TCR $\gamma$  is transcribed and spliced is shown in Figure 6. The nucleotide sequence of the TCR $\gamma$  transcript, as obtained from LNCaP, is shown in Table 1. The composite sequence is  $1020 \pm 3$  nucleotides long. It contains ~53 bases from the J $\gamma$ 1.2 gene segment, 519 bases of C $\gamma$ 1, followed by 448 bases of untranslated sequence containing a polyadenylation signal and poly(A) sequence at the 3' end.

**J. *In vitro* translation of the prostate-specific TCR $\gamma$  transcript.**

The prostate transcript has four translational initiation codons (ATG) in the original TCR $\gamma$  reading frame that are double underlined in Table 1. Calculated protein sizes for the four different start points are 12.8, 12.0, 7.2 and 3.2 kDa, respectively. To analyze the translational activity of the prostate transcript, *in vitro* transcription-coupled translation was carried out using full-length prostate TCR $\gamma$  cDNA. Two proteins of approximately 8 and 13 kDa were obtained, Figure 7 (lane 1). Negative control reactions did not yield any protein product.

## DISCUSSION

**Specific expression of TCR $\gamma$  transcripts in epithelial cells of the prostate.**

We identified expression of T-cell receptor  $\gamma$ -chain (TCR $\gamma$ ) mRNA in human prostate and have shown that it originates from epithelial cells of the prostate and not from infiltrating  $\gamma\delta$  T-lymphocytes. We also demonstrated that the T-cell receptor  $\delta$ -chain (TCR $\delta$ ) gene is silent in prostate. TCR $\gamma$  mRNA is expressed in epithelial cells within the acinar ducts of the prostate as well as in prostate cancer. Two TCR $\gamma$  transcripts of 1.1 kb and 2.8 kb are present in human prostate. They are different in size compared to the 1.5 kb TCR $\gamma$  transcript found in spleen, thymus and peripheral blood leukocytes. The TCR $\gamma\delta$  mRNA expression profile suggests that the transcription in prostate does not follow the usual pathway of  $\gamma\delta$  T-lymphocytes. The prostate TCR $\gamma$  expression was initially discovered by analysis of the publicly available EST database. Our results show

that EST clustering is a powerful tool to identify novel and unexpected gene expression. The prostate ESTs representing the TCR $\gamma$  transcript are all from cDNA libraries made from cells isolated by laser capture microdissection (Emmert-Buck *et al.*, *Science* 274, 998-1001 (1996)). The fact that the TCR $\gamma$  transcripts proved to originate from prostate epithelial cells and not from infiltrating  $\gamma\delta$  T-lymphocytes verifies that microdissection is a valuable technique to procure pure cell subpopulations from specific microscopic regions of tissues.

#### **The prostate TCR(JC) $\gamma$ transcript.**

The prostatic adenocarcinoma cell line, LNCaP, which was isolated from a lymph node metastasis (Horoszewicz *et al.*, *Cancer Res.* 43, 1809-1818 (1983)) expresses readily detectable levels of the 1.1 kb prostate-specific TCR $\gamma$  transcript. The expression in LNCaP cells shows that the transcript originates from epithelial cells and that it can be carried on during the development of a prostatic malignancy. The LNCaP transcript consists of ~53 bases of the J $\gamma$ 1.2 gene segment, the three C $\gamma$ 1 exons, untranslated sequence followed by poly(A) sequence. The prostate transcript is different from the mature T-lymphocyte transcript in that it lacks a V $\gamma$  gene segment and that it is initiated within the intronic sequence directly upstream of J $\gamma$ 1.2 (data not shown). The promoter driving the prostate TCR $\gamma$  transcript and its mechanism of activation in prostate epithelial cells are under investigation. The 2.8 kb prostate-specific TCR $\gamma$  transcript is very faint in LNCaP and the 5' RACE PCR experiment did not retrieve any product consistent with a 2.8 kb transcript. Therefore, the 2.8 kb transcript needs further study.

#### **Comparison with TCR (JC) $\gamma$ transcripts in T-lymphocytes.**

Many studies have shown that it is possible to detect TCR gene transcription prior to, or concomitant with, the onset of V(D)J rearrangement in hematopoietic cells (Wang *et al.*, *Mol. Immunol.* 33, 957-964 (1996); Shimamura, M., and Ohta, S., *Eur. J. Immunol.* 25, 1541-1546 (1995); Villey *et al.*, *Eur. J. Immunol.* 27, 1619-1625 (1997); Sikes *et al.*, *J. Immunol.* 161, 1399-1405 (1998)). The TCR $\gamma$  gene has been reported to be transcriptionally active in murine bone marrow-resident T-lymphocyte precursor cells with unrearranged  $\gamma$  loci, resulting in sterile TCR C $\gamma$  transcripts (Wang *et al.*, *Mol. Immunol.* 33, 957-964 (1996)). In addition, expression of unrearranged TCR V $\gamma$  transcripts have also been reported during ontogeny (Goldman *et al.* *J. Exp. Med.* 177, 729-739 (1993)). Sterile transcription of TCR and immunoglobulin gene segments has so

far been limited to cells from the lymphoid lineages (Lauzurica, P., and Krangel, M.S., *J. Exp. Med.* 179, 1913-1921 (1994)). Furthermore, activation of germ-line transcription at nearly all TCR and immunoglobulin loci temporally correlates with activation of locus recombination (Sikes *et al.*, *J. Immunol.* 161, 1399-1405 (1998); Goldman *et al.* *J. Exp. Med.* 177, 729-739 (1993); Lauzurica, P., and Krangel, M.S., *J. Exp. Med.* 179, 1913-1921 (1994); Sleckman *et al.*, *Annu. Rev. Immunol.* 14, 459-481 (1996)). We have shown, by independent experiments using genomic DNA and cDNA, that recombination has not taken place of the TCR $\gamma$  locus of prostate epithelial cells. Therefore, the expression of the TCR (JC) $\gamma$  transcript in prostate epithelium does not correlate with recombination and it may serve a different function than the sterile transcripts observed in T-lymphocyte precursor cells.

**Initial hypothesis of the possibility of a novel prostate-specific protein in the TCR $\gamma$  locus.**

The prostate TCR $\gamma$  transcript is highly expressed and we hypothesize that there is an underlying biologically important reason. The fact that VJ gene rearrangement has not taken place in the TCR $\gamma$  locus of prostate epithelial cells excludes the possibility that a mature TCR  $\gamma$ -chain protein is made. We also exclude the possibility that a TCR $\gamma$  constant domain protein is made without the TCR $\gamma$  variable domain, because no translational initiation codon (ATG) is found upstream of C $\gamma$ . In TCR  $\gamma$ -chain proteins a J $\gamma$  segment encodes 16-20 amino acids of the variable domain, while the major part of the variable domain is encoded by one of the V $\gamma$  segments. Unless the amino acids encoded by a J $\gamma$  segment are combined with amino acids encoded by a V $\gamma$  gene segment, they cannot function as a TCR in MHC recognition. This raised the possibility of a novel prostate-specific protein, encoded from within C $\gamma$ . Our initial hypothesis was that one of the ATG codons in the original TCR $\gamma$  reading frame initiates translation, although a different reading frame or a less frequently used initiation codon may be used.

The *in vitro* transcription-coupled translation experiment, using prostate TCR $\gamma$  cDNA revealed that the transcript was fully functional. Two proteins were obtained. The 13 kDa protein most likely originates from the first double underlined ATG in Figure 1, which yield a calculated protein size of 12.8 kDa (PS-TCR $\gamma$ -1). The 8 kDa protein most likely originates from the second double underlined ATG, calculated size of 7.2 kDa (PS-TCR $\gamma$ -2). These proteins were further explored in the studies reported in the next Example. In conclusion, the fact that prostate epithelial cells, or that

any non-lymphoid-derived cell type, express high level of a transcript from a gene that was thought to be exclusively expressed by cells from the lymphoid lineage, was a highly unexpected discovery.

## 5 EXAMPLE 2. DISCOVERY OF THE TCR $\gamma$ ALTERNATE READING FRAME PROTEIN

The previous Example demonstrated the unexpected discovery of TCR $\gamma$  transcript in prostate and prostate cancer cells, the in vitro translation of the transcript, and the initial hypothesis that the transcript resulted in the presence of a truncated form of  
10 TCR $\gamma$  chain in these cells. This Example sets forth the further unexpected discovery that the transcript in fact results in a previously unknown protein, now designated "TARP," expressed from an alternate reading frame. Even more unexpectedly, the studies reported below show that TARP is a nuclear protein, and is present in many breast cancer cells.

## 15 MATERIALS AND METHODS

**Primers.** TCR $\gamma$ -upATGmut#1 (5'-TTACAGATAAACAACCTTGATAC AGATGTTTCCCCCAAGCCC-3'); TCR $\gamma$ -upATGmut#2 (5'-GGGCTTGGGGGAAAC ATCTGTATCAAGTTGTTTATCTGTAA-3'); TCR $\gamma$ -upATGmut#3 (5'- GATAAACAACCTTGATGCAGATATTTCCCCCAAGCCC-3'); TCR $\gamma$ -upATGmut#4  
20 (5'-GGGCTTGGGGGAAATATCTGCATCAAGTTGTTTATC-3'); TCR $\gamma$ - upATGmut#5 (5'-GATAAACAACCTTGATACAGATATTTCCCCCAAGCCC-3'); TCR $\gamma$ -upATGmut#6 (5'-GGGCTTGGGGGAAATATCTGTATCAAGTTGTTTATC- 3'); TCR $\gamma$ -downATGmut#1 (5'-CCCAGGAGGGGAACACCATAAAGACTAAC GACACATAC-3'); TCR $\gamma$ -downATGmut#2 (5'-GTATGTGTCGTTAGTCTTT ATGGTGTTCCCCTCCTGGG-3'); TCR5.1 (5'-GATAAACAACCTTGATGCA  
25 GATGTTTCC-3'); TCR3.1 (5'-TTATGATTTCTCTCCATTGCAGCAG-3'); TCR $\gamma$ 1.2R (5'-AAGCTTTGTTCCGGGACCAAATAC); B-Actin Forward (5'- ATCTGGCACCACACCTTCTACAATGAGCTGCG-3'); B-Actin Reverse (5'- CTCATACTCCTGCTTGCTGATCCACATCTGC-3'). Primers were synthesized by  
30 Sigma-Genosys (The Woodlands, TX) and Lofstrand Labs Limited (Gaithersburg, MD).

**Constructs.** The *TARP* transcript cloned into pBluescript II SK(+) (Stratagene, La Jolla, CA) was described previously (Essand, M. *et al.*, *Proc. Natl. Acad. Sci. USA* 96:9287-9292 (1999)). This plasmid is referred to as pBSSK-TCR $\gamma$  in this

manuscript. pBSSK-TCR $\gamma$ mutATGup1, with the ATG at position 69 mutated to ATA, was constructed using the Quickchange Site-Directed Mutagenesis kit (Stratagene). The PCR reaction used TCR $\gamma$ -upATGmut#1 and TCR $\gamma$ -upATGmut#2 as primers and pBSSK-TCR $\gamma$  as template. pBSSK-TCR $\gamma$ mutATGup2, with the ATG at position 73 mutated to ATA, was constructed as above using TCR $\gamma$ -upATGmut#3 and TCR $\gamma$ -upATGmut#4 as primers and pBSSK-TCR $\gamma$  as template. pBSSK-TCR $\gamma$ mutATGup-both, with the ATGs at positions 69 and 73 mutated to ATA, was constructed as above using TCR $\gamma$ -upATGmut#5 and TCR $\gamma$ -upATGmut#6 as primers and pBSSK-TCR $\gamma$ mutATGup1 as template. pBSSK-TCR $\gamma$ mutATGdown, with the ATG at position 242 mutated to ATA, was constructed as above using TCR $\gamma$ -downATGmut#1 and TCR $\gamma$ -downATGmut#2 as primers and pBSSK-TCR $\gamma$  as template. pET-TCR $\gamma$  contains nucleotides 242-469 of the *TARP* transcript (Essand, M. *et al.*, *Proc. Natl. Acad. Sci. USA* 96:9287-9292 (1999)) subcloned into the pET23a vector (Novagen, Madison, WI). pET-TARP contains nucleotides 56-242 of the *TARP* transcript (Essand, M. *et al.*, *Proc. Natl. Acad. Sci. USA* 96:9287-9292 (1999)) subcloned into the pET23a vector. pVC4D-TARP contains nucleotides 69-242 of the *TARP* transcript (Essand, M. *et al.*, *Proc. Natl. Acad. Sci. USA* 96:9287-9292 (1999)) subcloned into the pVC4D vector (Bruggemann, E. P. *et al.*, *BioTechniques* 10:202-209 (1991)).

#### Reverse Transcription-Polymerase Chain Reaction (RT-PCR).

Isolation of poly(A) RNA was performed using the Micro-FastTrack™ 2.0 kit (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. 500 ng of poly(A) RNA or 5  $\mu$ g of total RNA were denatured for 2 minutes at 70°C in the presence of 50 pmol oligo-dT primer (Invitrogen). Single stranded cDNAs were prepared in a 10  $\mu$ l reaction containing 250  $\mu$ M dNTPs, 2 mM DTT, 8 U RNasin (Roche Molecular Biochemicals, Indianapolis, IN), 50 U Superscript II™ RT (Life Technologies, Rockville, MD) and incubated for 90 minutes at 42 °C. The samples were then diluted with 75  $\mu$ l 10 mM Tris-HCl [pH 7.5] and incubated at 72 °C for 10 minutes. 3  $\mu$ l of cDNA were used for PCR that contained 250  $\mu$ M dNTPs, 25 pmol of each respective primer, 1 unit AmpliTaq® DNA polymerase (Roche) and amplified for 35 cycles. Similar PCR conditions were used on the human breast RAPID-SCAN™ gene expression panel (OriGene Technologies, Rockville, MD). Primers TCR $\gamma$ J1.2R, TCR5.1 and TCR3.1 were used to detect the *TARP* transcript, while primers B-Actin Forward and B-Actin Reverse were used to detect the actin transcript.

**Northern Blot Hybridization.** Northern blot hybridization using 2 µg of poly(A) RNA was performed as described previously (Essand, M. *et al.*, *Proc. Natl. Acad. Sci. USA* 96:9287-9292 (1999)).

**In Vitro Transcription-Coupled Translation.** *In vitro* transcription-coupled translation reactions were described previously (Essand, M. *et al.*, *Proc. Natl. Acad. Sci. USA* 96:9287-9292 (1999)). pBSSK-TCR $\gamma$ , pBSSK-TCR $\gamma$ mutATGdown, pBSSK-TCR $\gamma$ mutATGup1, pBSSK-TCR $\gamma$ mutATGup2 and pBSSK-TCR $\gamma$ mutATGup- both were used as templates.

**Cell Culture.** LNCaP, PC3, MCF7, BT-474 and SK-BR-3 cells were maintained in RPMI-1640 medium (Quality Biological, Inc., Gaithersburg, MD) at 37 °C with 5% CO<sub>2</sub>. The medium contained 10% fetal bovine serum (FBS, Quality Biological, Inc.), 2 mM L-glutamine, 1 mM sodium pyruvate and penicillin/streptomycin. Hs57Bst cells were maintained in RPMI-1640 medium at 37 °C with 5% CO<sub>2</sub>. The medium contained 10% FBS, 30 ng/ml epidermal growth factor (EGF, Harlan, Cincinnati, OH), 2 mM L-glutamine, 1 mM sodium pyruvate and penicillin/streptomycin.

**Antibody Production.** Polyclonal  $\Delta$ PE-TARP antibodies were made as follows. pVC4D-TARP, which contains the entire TARP open reading frame fused to the C'-terminus of a catalytically inactive form of the *Pseudomonas* exotoxin ( $\Delta$ PE) (Bruggemann, E. P. *et al.*, *BioTechniques* 10:202-209 (1991)), was expressed in Epicurian Coli<sup>®</sup> BL21-CodonPlus<sup>™</sup> (DE3)-RIL cells (Stratagene). Preparation of inclusion bodies and rabbit immunization were described previously (Brinkmann, U. *et al.*, *Proc. Natl. Acad. Sci. USA* 88:8616-8620 (1991)). The antiserum was purified using the ImmunoPure<sup>®</sup> IgG (Protein A) Purification Kit according to the manufacturer's instructions (Pierce, Rockford, IL).

TCR $\gamma$  antibodies were made as described above using pET-TCR $\gamma$ , an expression plasmid containing the extracellular domain of TCR $\gamma$  fused to a C'-terminal six-histidine tag. Prior to immunization, the histidine-tagged TCR $\gamma$  protein was purified using a Ni-NTA agarose column according to the manufacturer's instructions (QIAGEN, Valencia, CA).

**Preparation of Cell Extracts.** Whole cell protein extracts were prepared as follows. 5 x 10<sup>6</sup> growing cells from each respective cell line were harvested and resuspended in 1X RIPA buffer containing proteinase inhibitors (50 mM Tris-HCl [pH 7.5], 150 mM NaCl, 1 mM EDTA, 0.1% TritonX-100, 1 mM PMSF, 1 µg/ml aprotinin, 1

µg/ml leupeptin). The extracts were sonicated briefly and clarified by centrifugation. Protein concentrations were determined using the Coomassie® Plus Protein Assay reagent according to the manufacturer's instructions (Pierce). Protein extracts from prostate tissue were prepared by grinding 0.5 g of prostate cancer tissue frozen at -80 °C into a fine powder using a cold mortar and pestle. The powdered tissue was collected, resuspended in 1X RIPA and processed as described above.

Nuclear, membrane and cytoplasmic extracts from prostate and breast cell lines were prepared based on protocols previously published (Dignam, J. D. *et al.*, *Nucleic Acids Res.* 11:1475-1489 (1983); Sladek, F. M. *et al.*, *Genes Dev.* 4:2353-2365 (1990)).

**Western Blot Analysis.** 20 or 40 µg of protein extract, 1 µg of recombinant His-TARP or 100 ng of recombinant His-TCRγ were run on a 16.5% Tris-Tricine gel (BIO-RAD, Hercules, CA) and transferred to a 0.2µm Immobilon-P™ PVDF membrane (BIO-RAD) in transfer buffer (25 mM Tris, 192 mM glycine, 20% (v/v) methanol, pH 8.3) at 4 °C for 4 hours at 30 V. Filters were probed with either 10 µg/ml ΔPE-TARP antiserum or 1 µg/ml TCRγ antiserum and their respective signals were detected using a chemiluminescence western blotting kit according to the manufacturer's instructions (Roche).

**TARP is a nuclear protein expressed in prostate cancer cells.** To determine whether TARP or TCRγ exists in prostate cancer cells, we generated antibodies against both proteins and performed western blots on different prostate cancer cell extracts. As shown in Figure 3A (top panel), TARP was detected in the prostate cancer LNCaP cell line and a prostate cancer tumor extract. The 7 kDa band comigrates with the recombinant His-TARP suggesting that the product detected in the LNCaP and cancer extracts is TARP. Previously, we demonstrated that the prostate-specific TCRγ transcript is not expressed in the prostate cancer PC3 cell line (Essand, M. *et al.*, *Proc. Natl. Acad. Sci. USA* 96:9287-9292 (1999)). Therefore, we used PC3 cell extracts as a negative control and demonstrated that the 7 kDa band was absent in these extracts (Figure 3A, top panel). Importantly, no 7 kDa bands were detected when the pre-bleed antiserum or an antiserum against the *Pseudomonas* exotoxin (PE, see Materials and Methods) was used (data not shown). TCRγ was not detected in any of these extracts even though the recombinant protein showed a very strong signal with the antibody employed (Figure 3A,

bottom panel). These data indicate that the prostate-specific *TCR $\gamma$*  transcript encodes TARP.

To determine the cellular localization of TARP, we prepared nuclear, cytoplasmic and membrane fractions from LNCaP cells. As shown in Figure 3B, TARP was detected in the nucleus and not in the cytoplasm or membrane fraction. Similar results were obtained using nuclei purified by fractionating the cell extracts through a sucrose cushion (Sladek, F. M. *et al.*, *Genes Dev.* 4:2353-2365 (1990)) (data not shown).

**The *TARP* transcript is expressed in breast cells.** Previously, we reported that the *TCR $\gamma$*  EST cluster also contains some ESTs from brain libraries (Vasmatzis, G. *et al.*, *Proc. Natl. Acad. Sci. USA* 95:300-304 (1998)). After this initial report, additional ESTs have been deposited into the database and the cluster now contains ESTs from breast, colon, kidney and gastric libraries as well. To determine whether the existence of these ESTs indicates the expression of the *TARP* transcript in these cells or whether it may due to the presence of infiltrating  $\gamma\delta$  T-lymphocytes when these libraries were made, we performed RT-PCR on various cell lines to test for the presence of the *TARP* transcript. As shown in Figure 4A, expression of the *TARP* transcript was detected in the breast cell lines MCF7, BT-474, SK-BR-3 and CRL-1897. No signals were detected in the neuroblastoma cell line A172, glioblastoma cell line IMR32, colon cell line COLO 205, gastric cell line KATO III or kidney cell lines COS7 and 293 (Figure 4A and data not shown). To determine whether the *TARP* transcript is expressed in human breast tissues in addition to cell lines, we tested 12 different normal breast and 12 different breast cancer cDNAs using a RAPID-SCAN<sup>TM</sup> panel (OriGene Technologies, Rockville, MD). *TARP* mRNA was shown to be abundant in some of the breast cancer samples (Figure 4B, top panel) while barely detectable in the normal breast samples after 35 rounds of PCR (data not shown). Significantly, no signals were detected in reactions lacking cDNA. *Actin* was used to show that similar amounts of cDNA were present in each lane (bottom panel). The weak signals in the normal breast samples correlate well with the lack of *TARP* signal shown in Figures 4A and 5 for the Hs57Bst cell line, a breast cell line derived from normal breast tissue. These results suggest that expression of the *TARP* transcript in the breast is increased after oncogenic transformation. However, more studies are needed before any definitive conclusions can be made.



To determine whether the *TARP* transcript observed in the breast cell lines is the same as the transcript found in the prostate cell line, we performed RT-PCR using primers against different regions of the *TARP* transcript. As shown in Figure 5A, the *TARP* transcript in prostate contains a portion of the Jy1.2 gene segment, three C $\gamma$ 1 exons and some untranslated sequence followed by a poly(A) tail (7). Primer set 1 and 3 amplifies the entire *TARP* transcript (Figure 5B, top panel) while primer set 2 and 3 amplifies the C $\gamma$ 1 region only (Figure 5B, middle panel). As shown in Figure 5B, similar-sized bands were detected in three breast cell lines (MCR7, BT-474 and SK-BR-3) as compared to the prostate cell line (LNCaP) using either primer set. Importantly, no signals were detected in the reactions lacking cDNA (dH<sub>2</sub>O) and similar amounts of cDNA were used as demonstrated by the *actin* control (Figure 5B, bottom panel). These data indicate that the *TARP* transcript found in the breast cell lines is the same as the transcript found in the prostate cell line. To further support this conclusion, we analyzed the *TARP* transcript sizes from each cell line by a northern blot. Previously, we showed that 1100 and 2800 nucleotide transcripts exist in LNCaP cells, with the 1100 nucleotide transcript being the predominant form (Essand, M. *et al.*, *Proc. Natl. Acad. Sci. USA* 96:9287-9292 (1999)). As shown in Figure 5C, similar-sized *TARP* transcripts were found in three breast cell lines (MCF7, BT-474 and SK-BR-3) as compared to the prostate cell line (LNCaP), although at a weaker intensity. Therefore, we conclude that *TARP* mRNA is expressed in prostate and breast cancer cells.

To determine whether TARP protein exists in the breast cancer cell lines, we performed a western blot with breast cancer nuclear extracts using an antibody against TARP. As shown in Figure 6 (top panel), TARP reactive bands were detected in MCF7, BT-474 and SK-BR-3 cells. TARP was not detected in the membrane or cytoplasmic fractions in these breast cancer lines (data not shown). Importantly, TARP is the protein product encoded by the *TARP* transcript in the breast cell lines because TCR $\gamma$  was not detected in any of these nuclear extracts even though the recombinant protein showed a very strong signal with the antibody employed (Figure 6, bottom panel). These data indicate that TARP also exists in breast cancer cells.

We report the identification of a 7 kDa nuclear protein encoded by a specific transcript derived from the *TCR $\gamma$*  locus expressed in prostate and breast cancer cells. Because the protein is encoded from a reading frame different from TCR $\gamma$ , we name it TARP for TCR $\gamma$  Alternate Reading frame Protein. Besides being translated from

an alternate reading frame of a transcript originating within an intron of the *TCR $\gamma$*  locus, TARP has two other unusual features. First, it is surprising to find such a small peptide in the cell because most are usually secreted. Second, TARP lacks a good Kozak sequence (Kozak, M. *Cell* 44:283-92 (1986)). Because the *TCR $\gamma$*  reading frame contains a good Kozak sequence, we initially hypothesized that a truncated *TCR $\gamma$*  protein was encoded. However, as shown in Figure 3, our initial hypothesis was incorrect. It is of interest that the *in vitro* translation results indicate a preference for the TARP protein and that either ATG in the TARP reading frame can be used to initiate protein synthesis. Protein sequencing will be needed to determine which ATG is used to initiate TARP protein synthesis.

A very interesting feature of the TARP protein sequence is that it contains five leucines in heptad repeats, suggesting that TARP may contain a leucine zipper dimerization motif (Figure 7A). For this to be true, TARP must contain an amphipathic helix. One indication that TARP may contain an amphipathic helix is that serine and proline residues, residues believed to serve as a helix initiator, are found immediately before the first leucine repeat. Second, many charged amino acids are found within the heptad repeats thereby giving the helix an amphipathic nature and potentially serving as salt bridges with other helices. Even though the presence of leucines in heptad repeats is a good indication of a leucine zipper motif, there are proteins identified containing five leucines in heptad repeats that are not considered leucine zipper proteins. For example, the crystal structures for karyopherin (Chook, Y. M. *et al.*, *Nature* 399:230-237 (1999)), *B. stearothermophilus* pyrimidine nucleoside phosphorylase (Pugmire, M. J. *et al.*, *Structure* 6:1467-1479 (1998)) and *T. thermophilus* phenylalanyl-tRNA synthetase (Mosyak, L. *et al.*, *Nat. Struct. Biol.* 2:537-547 (1995)) have shown that these proteins do not contain  $\alpha$ -helical structures in the region where the sequence contains five leucines in heptad repeats. Interaction and structure studies are needed to determine the significance of the leucine repeats found in TARP.

Another unusual feature of the TARP amino acid sequence is that a region of basic amino acids follows the potential leucine zipper motif (Figure 7A), suggesting a possible DNA-binding motif. However, the orientation of the basic region is rather unique in that it follows the leucine repeats rather than precedes them. Most leucine zipper proteins that bind DNA have the basic region before the leucine repeats (for a review, see (Chook, Y. M. *et al.*, *Nature* 399:230-237 (1999))). The basic region in

TARP may only be functioning as a nuclear localization signal, but the fact that TARP is a nuclear protein strengthens the hypothesis that TARP may bind DNA. Functional studies are needed before any definitive conclusions can be made.

To determine if TARP shares homology with any known proteins, we performed a protein BLAST search against GenBank. This search indicated that the amino acid sequence of TARP shares some homology to *Dictyostelium discoideum* Tup1 (GenBank accession no. AAC29438) and *Saccharomyces cerevisiae* Tup1 (Williams, F. E. *et al.*, *Mol. Cell. Biol.* 10:6500-6511 (1990)) (Figure 7C). Yeast Tup1 is normally found in a complex with Cyc8(Ssn6) and is required for transcriptional repression of genes that are regulated by glucose, oxygen and DNA damage (Tzamarias, D. *et al.*, *Genes Dev.* 9:821-831 (1995)). Neither Cyc8(Ssn6) nor Tup1 binds DNA, but each acts as a part of a corepressor complex through interactions with specific DNA-binding proteins such as  $\alpha 2$ , Mig1, Rox1 and a1 (Tzamarias, D. *et al.*, *Genes Dev.* 9:821-831 (1995)). The C'-terminal half of Tup1 contains six repeats of a 43-amino acid sequence rich in aspartate and tryptophan, known as WD-40 or  $\beta$ -transducin repeats (Williams, F. E. *et al.*, *Mol. Cell. Biol.* 10:6500-6511 (1990); Fong, H. K. *et al.*, *Proc. Natl. Acad. Sci. USA* 83:2162-2166 (1986)). WD-40 repeats have been identified in many proteins and play a role in protein-protein interactions. Importantly, Tup1 has been shown to interact with  $\alpha 2$  through two of its WD-40 repeats (Komachi, K. *et al.*, *Genes Dev.* 8:2857-2867 (1994)). It is interesting to note that TARP shares homology with the fifth WD-40 repeat of Tup1 (Figure 7C). Because TARP is a nuclear protein, its homology with Tup1 suggests that TARP may be a member of a functional nuclear protein complex involved in transcriptional regulation. Therefore, it is necessary to identify TARP-interacting proteins in order to determine its function.

The TARP antibody recognizes a doublet in prostate and breast nuclear extracts (Figure 6A). The faster 7 kDa band comigrates with the His-TARP recombinant protein, while the weaker band runs at a larger molecular weight. One possible explanation for the 9 kDa band is post-translational modifications. To determine if TARP contains any known post-translational modification sites, we analyzed the TARP amino acid sequence using the PROSITE program of the *Swiss Institute of Bioinformatics* ExPASy proteomics server (<http://www.expasy.ch>) (Appel, R. D. *et al.*, *Trends Biochem. Sci.* 19:248-260 (1994); Hofmann, K. *et al.*, *Nucleic Acids Res.* 27:215-219 (1999)). As shown in Figure 7A, many potential phosphorylation sites were found including cAMP- and cGMP-dependent protein kinase phosphorylation sites (RRAT and RRGTT) and

protein kinase C phosphorylation sites (SSR and SRR). Phosphorylation has been shown in many cases to cause a protein to run at a larger apparent molecular weight on an SDS-PAGE gel. If this is the case, the results from Figure 6 may indicate that the unmodified form is prevalent in LNCaP cells and that only the phosphorylated form is present in MCF7 and SK-BR-3 cells. Additional experiments are clearly needed to determine the true nature of the 9 kDa band and whether TARP is post-translationally modified when expressed in prostate and breast cancer cells.

We report here the expression of TARP mRNA and protein in breast cancer cells. Our initial studies of the *TARP* transcript did not reveal *TARP* expression in the breast (Essand, M. *et al.*, *Proc. Natl. Acad. Sci. USA* 96:9287-9292 (1999)). One possible explanation is that *TARP* is expressed at low levels in the normal breast and is difficult to detect. As described in the Results section, very weak signals were detected in a PCR analysis of normal breast samples as compared to the strong signals detected in the cancer samples. Therefore, the presence of *TARP* in breast cancer cells may indicate that *TARP* expression is induced after the oncogenic transformation of breast cells. In addition, the existence of *TARP* in breast cancer cells may indicate that *TARP* is regulated by estrogen. This hypothesis is strengthened by the identification of an element within the intronic promoter of *TARP* that combines an androgen response element (ARE) with an estrogen response element (ERE). This hybrid element consists of two half-sites specific to the ARE at the 5' end and to the ERE at the 3' end [(Zilliacus, J. *et al.*, *Mol. Endocrinol.* 9:389-400 (1995)) and unpublished data]. Additional experiments are needed to determine if estrogen regulates *TARP*. There are instances, however, where mutant AREs cause the expression of certain prostate-specific genes in breast tumors. For example, *prostate specific antigen* (PSA) has been shown to be expressed in breast tumors (Majumdar, S. *et al.*, *Br. J. Cancer* 79:1594-1602 (1999)). Molecular analysis of the aberrant expression of PSA lead to the discovery of a single point mutation in one of the AREs found within the PSA promoter. It is believed that this mutation leads to the loss of androgen-regulated PSA expression in breast tumors (Majumdar, S. *et al.*, *Br. J. Cancer* 79:1594-1602 (1999)). It is unclear at this time whether a similar mutation in the *TARP* promoter occurs in the three breast cell lines tested.

The prostate is dependent on androgens for maintenance of its structure and function. When prostate cells become malignant, they often lose their androgen dependence. In this study, we used two prostate cell lines that differ in their dependence on androgen for growth: LNCaP and PC3 cells. The androgen receptor is present in the

androgen-dependent LNCaP cell line, but is absent in the androgen-independent PC3 cell line (Tilley, W. D. *et al.*, *Cancer Res.* 50:5382-5386 (1990)). As shown in Figure 3, *TARP* is expressed in LNCaP cells but not in PC3 cells. This result suggests that *TARP* expression may be regulated by androgen stimulation. The identification of an ARE-like element within the *TARP* promoter strengthens the idea that *TARP* is induced by androgens. Experiments are currently being done to determine whether androgens induce *TARP* mRNA expression. Expression in LNCaP cell but not in PC3 cells may indicate that *TARP* is important in regulating androgen-dependent responses.

10

The present invention provides novel materials and methods relating, *inter alia*, to prostate cells, prostate cancer, and breast cells and breast cancer. While specific examples have been provided, the above description is illustrative and not restrictive. Many variations of the invention will become apparent to those skilled in the art upon review of this specification. The scope of the invention should, therefore, be determined not with reference to the above description, but instead should be determined with reference to the appended claims along with their full scope of equivalents.

All publications and patent documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication or patent document were so individually denoted. By their citation of various references in this document Applicants do not admit that any particular reference is "prior art" to their invention.

20